Ca' Foscari
University
of Venice

Doctoral Programme in Science and Technology of Bio and Nanomaterials

Doctoral Thesis

Deep Learning Approaches to the Automated Interpretation of One-Dimensional ^{1}H NMR Spectra

Supervisors
Ch. Prof. Guido Caldarelli

Ch. Prof. Alessandro Scarso
Ch. Prof. Dirk Wilhelm

PhD Candidate
Giulia Fischetti
Matriculation Number 956653

Academic Year
2025 / 2026

Declaration on the Use of Generative AI

I hereby declare that the generative AI tool Gemini Pro (accessed via the institutional license provided by Ca' Foscari University of Venice) was utilized during the preparation of this thesis. Its use was strictly limited to the revision phase to improve grammatical accuracy, syntax, and overall text clarity. The core ideas, intellectual content, data analysis, and conclusions presented in this work are entirely my own.

Acknowledgements

I am immensely grateful to have had the opportunity to conduct my doctoral research within the lively, stimulating, and international context bridging Ca' Foscari University of Venice and the Zurich University of Applied Sciences (ZHAW). Working on applied projects alongside industrial partners allowed my research to have a tangible, real-world impact, making this journey incredibly rewarding right from the start.

My deepest gratitude goes to my supervisors, *Prof. Guido Caldarelli* and *Prof. Alessandro Scarso* at Ca' Foscari, and *Prof. Dirk Wilhelm* at ZHAW, for their invaluable guidance, for helping me develop my ideas, and for always granting me the intellectual space to grow into them.

I would also like to extend my thanks to our industrial collaborators. My gratitude goes to the Bruker BioSpin Group in Switzerland, and especially to *Simon Bruderer*, for the tight exchange and their experienced perspective throughout our cooperation. I am equally thankful to my current colleagues at NexMR AG, *Félix Torres* and *Mathias Bütikofer*, for their constant hands-on availability and insights.

I am also deeply grateful to Prof. Rudolph Fuchslin for providing his open support and scientific esteem.

At ZHAW, I was fortunate to be surrounded by fantastic colleagues. A special thanks goes to *Nicolas*, for his unconditional support on both a professional and deeply human level. I am also grateful to *Andreas* for strenghtening my confidence as I transitioned from a student into a more independent researcher.

My warm acknowledge goes to *Diego*, *Felix*, *Joram*, and *Peter*, with whom I've been sharing the daily struggles of research life and whose company makes even the hardest days lighter. Furthermore, I am thankful for the connections made along the way at international conferences, specifically with *Anne-Marie*, *Laura*, *Maik*, and *Jennifer*, who made me experience firsthand how doing science is truly a community endeavour.

Life has a beautiful way of aligning things, and I am uniquely grateful for three unexpected circumstances since arriving in Switzerland. Had I already learned German, I would have missed the chance to get to know *Johanna*. Had I known how to ski before moving here, I would never have met *Gloria*. And had a new PhD position not unexpectedly opened up this year, *Kinga* would not have crossed my path. These three exceptional individuals have truly become my anchors as I navigate my life abroad.

Across the distance, *Lorenzo* has known me since my teenage years and has always felt close, no matter how many kilometres separate us, a rare and precious kind of friendship. Finally, and most profoundly, I thank my family: *my mum*, *my dad*, and *Gabriele*, who have fought for my happiness at every turn, given me the strength to continue, and without whom none of this would mean very much at all.

Abstract

One-dimensional proton Nuclear Magnetic Resonance (^{1}H NMR) spectroscopy is broadly adopted in chemical sciences for structure characterization due to its rapid acquisition and experimental simplicity. While significant technical progress has enabled high-throughput data collection, the utility of these datasets is currently limited by a much slower process of manual spectral analysis. The reliance on expert annotation is not only labor-intensive but introduces variability that can compromise the reproducibility of quantitative results. These issues are further complicated by physical constraints such as spectral overlap, solvent interference, and signal distortions, which often obscure critical molecular information.

This doctoral thesis investigates the encoding of molecular information in NMR signals and develops efficient, uncertainty-aware algorithms for automated interpretation. The central contribution is MuSe Net, a supervised probabilistic deep learning framework designed for automated signal detection and splitting pattern classification. By treating multiplet analysis as a one-dimensional segmentation task, MuSe Net provides point-level predictions that include both class labels and confidence scores. This uncertainty metric plays a critical role in the analysis, as it assists in the identification of overlapping and composite signals where spectral patterns deviate from first-order multiplet classes. To support the training of our model, a computationally efficient strategy was developed to generate realistic synthetic spectra based on specific multiplet phenotypes, bypassing the need for large-scale experimental spectra acquisition or complex quantum-mechanical simulations.

The research further addresses the integration and correlation of information across varying experimental conditions, specifically within the framework of fragment-based drug discovery with photo-Chemically Induced Dynamic Nuclear Polarization (photo-CIDNP) ^{1}H NMR. A dedicated computational pipeline was implemented to automatically detect hyperpolarized signals and evaluate signal attenuation between ligand-only and ligand–protein spectra. By employing a Bayesian assignment algorithm, estimated quenching values are mapped onto the molecular structure to identify potential binders. To ensure robustness, a specific processing protocol was developed to enforce a consistent absorptive character across spectra, utilizing symmetric peak shapes in the magnitude spectrum to improve automated detection.

Realized in partnership with Bruker Switzerland AG and NexMR AG, the algorithms presented in this work aim at bridging the gap between signal processing theory and analytical practice. This research provides a formalized alternative to manual annotation, accelerating NMR workflows and increasing the accessibility of high-quality spectral analysis in both academic and industrial settings.

Contents

Chapter 1

Introduction

Nuclear spin is the intrinsic quantum property underlying Nuclear Magnetic Resonance (NMR) spectroscopy. It gives rise to a magnetic dipole moment, which renders the nucleus responsive to an external magnetic field. The magnetic moment of the proton was first measured in 1933 by Stern [1]. Shortly thereafter, Isidor Rabi [2] demonstrated that nuclei placed in a strong static magnetic field undergo resonant transitions when exposed to a transverse oscillating field at a characteristic frequency determined by the local magnetic environment, establishing the fundamental principle of magnetic resonance. By 1945, two independent groups, Purcell and Pound [3], and Bloch, Hansen and Packard [4], had detected radio-frequency signals emitted by nuclear spins in bulk matter, effectively realizing the first experimental implementations of NMR spectroscopy. Two decades later, Richard Ernst [5] revolutionized the field by introducing Fourier transform NMR (FT-NMR), replacing slow continuous-wave excitation with short radio-frequency pulses to dramatically enhance both the sensitivity and resolution of the resulting spectra.

A one-dimensional NMR spectrum consists of resonance signals distributed along the frequency axis, ranging from isolated lines, singlets, to pairs of lines, doublets, and more complex groups of peaks collectively termed multiplets. The position of a multiplet along the frequency scale, the chemical shift, its integrated intensity, the internal splitting pattern due to spin-spin coupling jointly encode structural information about the relative abundance of nuclei, their local electronic environments, and the connectivity arrangement within the molecule.

Since the first experiments, the NMR field has developed rapidly and extensively, leading to a proliferation of methodologies and pulse sequence designs of increasing complexity and sophistication. Contemporary NMR experiments are capable of generating not only one-dimensional absorption spectra, but also multidimensional datasets in which correlations between nuclear spins encode detailed structural and dynamical information about the molecule under investigation [6].

Despite these advances, one-dimensional ^{1}H NMR spectroscopy remains one of the most widely employed NMR modalities. Its experimental implementation is comparatively simple, requires minimal specialized preparation beyond standard sample handling procedures, and permits rapid data acquisition. These characteristics render one-dimensional ^{1}H NMR a practical and broadly accessible method for the routine characterization of organic, inorganic, and biomolecular systems across diverse areas of scientific research. Within analytical and organic chemistry, one-dimensional ^{1}H NMR spectroscopy is routinely employed in the elucidation of newly isolated natural and synthetic reaction products [7], where spectral features provide constraints on molecular connectivity and

functional group composition. Beyond connectivity analysis, ^{1}H NMR enables the investigation of conformational equilibria through the sensitivity of scalar couplings and chemical shifts to molecular geometry. Furthermore, the technique permits the characterization of kinetic processes and dynamic exchange phenomena on timescales comparable to the NMR frequency differences, as manifested in line broadening, coalescence behavior, and temperature-dependent spectral changes. Owing to its quantitative nature, where under appropriately controlled acquisition conditions, signal integrals are proportional to molar concentration, NMR enables direct mixture analysis. This property is widely applied in metabolomics [8], where complex biofluids such as serum, plasma, and urine are profiled to quantify metabolites. Such analyses support biomarker discovery and disease classification, and facilitate the investigation of metabolic dynamics and pathway compartmentation within systems-level biochemical studies [9]. NMR spectroscopy plays also a significant role in ligand-based drug discovery, where it provides direct information on the nature and strength of molecular interactions and enables the identification of weak-binding ligands [10, 11, 12, 13]. In clinical and biomedical contexts, magnetic resonance techniques are employed for diagnostic purposes, including the detection and monitoring of pathological conditions such as tumors, hematomas, and multiple sclerosis [14]. NMR further contributes to materials science, for example in the study of lithium-ion batteries, where it provides insight into ion transport, electrode structure, and degradation mechanisms [15]. In environmental research, NMR is applied to the characterization of organic matter in soils, natural waters, and atmospheric samples, enabling the investigation of pollution effects and ecosystem changes [16, 17]. Additional applications include food science and agriculture, where NMR is used for authenticity assessment and adulteration testing, quality control of food products and beverages [18], and the analysis of plant tissues, soil components, and complex biomolecular mixtures relevant to crop [19] and nutritional studies [20].

The analytical potential of one-dimensional ^{1}H NMR spectroscopy is, however, limited by the typically time-consuming and labor-intensive nature of its data analysis. Conventional workflows involve multiple sequential steps, which are often performed manually or semi-manually and require comprehensive domain expertise. After data acquisition and Fourier transformation, the resulting complex spectrum is subjected to a series of pre-processing operations designed to correct phase and baseline distortions [21]. In conventional practice, once an acceptable phase correction has been achieved, the imaginary component is typically discarded, and subsequent analysis is performed exclusively on the real part of the spectrum, which is more readily interpretable in terms of absorptive line shapes. The initial step of spectral analysis consists in the identification and localization of individual resonance signals, commonly referred to as peak picking. This procedure is typically accompanied by the estimation of relevant peak characteristics, including amplitude, linewidth, and line shape parameters, which are subsequently used for quantitative analysis and structural interpretation. Identified resonances are subsequently grouped into multiplets and interpreted as signals arising from the same nucleus or from sets of magnetically equivalent nuclei. This procedure enables the analysis of splitting patterns and the determination of scalar spin–spin coupling constants, which encode information on through-bond connectivity. Quantitative information is obtained by integrating the signal intensity over the corresponding spectral regions, and estimating the relative abundances of the contributing nuclei, thereby permitting the inference of the underlying spin system. By combining information on spin connectivity with constraints derived from chemical shifts and local electronic environments, it becomes possible to

assign nuclei to their corresponding resonances and to formulate a structural proposal for the molecule under investigation. In practice, this process of structure elucidation often yields multiple plausible candidates. Discrimination among them typically requires additional NMR experiments, such as ^{13}C or two-dimensional correlation spectra, as well as comparison of the experimental data with spectra simulated from candidate structures (structure verification).

In addition to its labor-intensive character, one-dimensional ^{1}H NMR data analysis is further complicated by the presence of signals that do not originate from the compound of interest. These include experimental artefacts, spurious resonances, and residual solvent signals, all of which may confound spectral interpretation and bias quantitative estimates derived from signal integration. Intense solvent resonances are particularly problematic, as they may induce baseline distortions that hinder reliable pre-processing across the entire spectral range. Moreover, their shoulders and tails can mask nearby low-intensity signals, thereby obscuring relevant structural information. Similar difficulties arise in the presence of broad resonances generated by exchangeable nuclei, such as protons undergoing chemical exchange with solvent protons or deuterons on the NMR timescale. These exchange processes lead to line broadening and reduced spectral resolution, further complicating peak identification and quantitative analysis. Moreover, resonances arising from different nuclei may occur at very similar chemical shifts, leading to overlapping multiplets, i.e., superpositions of individual spectral components. Such overlap hampers the reliable analysis of splitting patterns and complicates the accurate determination of scalar spin–spin coupling constants. In addition, the quantitative estimation of relative abundances through signal integration becomes less reliable when spectral regions cannot be unambiguously attributed to a single group of equivalent spins. Consequently, in complex spectra the expert interpretation may not be univocal, leading to alternative, yet internally consistent, structural hypotheses, and the analysis can remain inconclusive. Even in cases where the spectrum is comparatively clear, manually annotated data often lack uniformity. Small variations in peak localization or in the definition of integration boundaries introduce intrinsic variability, which may propagate and become amplified in subsequent quantitative or comparative analyses.

The development of reliable automated procedures for the extraction of structural and quantitative information from one-dimensional ^{1}H NMR spectra has long represented a central objective in the evolution of the field. The motivation is both practical and methodological. By formalizing and systematizing the interpretative process, automation would provide several advantages. It would enable rapid and effective analysis, thereby substantially reducing the time required for spectral interpretation, an aspect of particular relevance in high-throughput settings. It could also reduce the need to employ more resource-intensive experimental techniques when the essential structural information is already contained in one-dimensional data. Standardized computational workflows would promote annotation consistency across laboratories and enhance the reproducibility of results within the scientific community. The availability of reliable automated tools would further lower the barrier to entry for non-specialists, facilitating the adoption of NMR spectroscopy in applied research contexts where user-independent and scalable analysis is required.

Early efforts toward automation relied primarily on rule-based algorithms designed to emulate manual inspection procedures. These approaches exploited local regularities of resonance signals, including the identification of local extrema, the analysis of spectral derivatives, the application of intensity or noise thresholds, and the use of geometric or

symmetry properties characteristic of resonance line shapes [22, 23]. In order to improve robustness in the presence of noise and to reduce spurious detections, several methods operated on transformed representations of the spectrum. Examples include wavelet-based decompositions [22], non-negative matrix factorization [24], singular value decomposition [25], and maximum-entropy processing [26, 23, 27, 28]. These techniques aimed to enhance signal-to-noise separation or to provide compact representations that facilitate peak detection under non-ideal experimental conditions. A complementary research direction addressed the problem of individual peak deconvolution through explicit parametric modeling of spectral line shapes. In this framework, resonances were represented as Lorentzian, Gaussian, or Voigt profiles, and their parameters were estimated either directly from the spectrum with the support of the derivative information [29, 30], or modeling the spectrum as a convolution of a peak-shape function with a sparse distribution of sources modeled as delta functions [31]. For multiplet deconvolution, clusters of resonances were modeled as superpositions of parametric line shapes whose relative frequency offsets and intensities encode scalar coupling constants [32, 33]. The analysis of multiplets has concentrated primarily on the structural organization of signal splitting. Golotvin and Chertkov [34] formulated multiplet interpretation as a pattern-recognition problem, in which extracted spectral features were used to classify resonance patterns. Hoye and co-workers [35, 36] instead proposed a hierarchical decomposition framework governed by scalar coupling relationships, subsequently refined through the incorporation of symmetry relations and amplitude constraints [37, 38]. A further line of work approached multiplet identification as a spectrum-wide organization task, clustering individually detected peaks into multiplet groups rather than treating resonances as purely local entities [39].

These traditional parametric and rule-based approaches, although effective in many practically relevant scenarios and still integrated into expert-driven workflows, critically depend on the specification of numerous analysis parameters. These include noise and intensity thresholds for peak detection, assumptions regarding line-shape models, and prior estimates of multiplet boundaries, peak counts, and coupling constants. Their performance is therefore closely tied to user-defined settings and to the validity of underlying modeling assumptions. In recent years, the emergence of artificial intelligence has profoundly changed the prevailing algorithmic paradigm, shifting attention from explicitly modeled representations toward data-driven methodologies. Across diverse domains, machine learning techniques have demonstrated broad applicability, ranging from regression and object detection to speech recognition and advanced natural language processing, including the development of autonomous agent-based systems. In chemistry, a particularly prominent milestone was the development of the *AlphaFold* system [40], which predicts three-dimensional protein structures directly from amino acid sequences with accuracy approaching experimental resolution. This transformation toward learning-based inference has also influenced the field of NMR spectroscopy [41, 42, 43, 44], where data-driven approaches are increasingly applied across the entire analysis pipeline: from signal reconstruction and denoising, to spectral interpretation, and ultimately to fully automated structure verification and elucidation.

Concerning NMR signal analysis, several deep learning approaches have been proposed for automated peak detection in two-dimensional spectra. These methods either operate on cropped spectral patches, or analyze the two frequency axes separately through convolutional or residual architectures designed to capture local spectral features [45, 46, 47]. More recent physics-informed architectures leverage the specific point spread function of

the sampling schedule to map experimental data directly to peak probability distributions [48]. While many of these networks were developed primarily for multidimensional protein NMR, a distinct line of work emerged in the context of our collaboration with Bruker Switzerland AG, a major spectrometer manufacturer, focusing on one-dimensional NMR of small molecules. In this study [49], peak detection and parameter estimation were formulated as a joint classification–regression problem, enabling the resolution of overlapping resonances in complex spectra. The resulting method demonstrated sufficient robustness and practical utility to be integrated into Bruker's commercial analysis software TopSpin, where it is available as part of the standard processing toolchain.

The analysis of splitting patterns is generally considered a more robust source of structural information than chemical shifts alone, as scalar coupling constants exhibit comparatively limited variability across experimental conditions [6]. Compared to peak detection and general spectral processing, fewer studies have addressed the automated extraction of multiplicities and coupling constants. Much of the existing work considers the inverse problem, namely the prediction of coupling constants from molecular structure. In such approaches, scalar couplings are estimated from structural descriptors encoding bond characteristics, electronic properties, and geometric parameters relevant to through-bond interactions [50, 51]. However, in many practically relevant applications the primary goal is to extract splitting-pattern information directly from experimental spectra in order to reconstruct an unknown molecular structure.

The first part of the present doctoral work is devoted to the development of robust and reliable methods for the automatic identification and characterization of multiplet splitting patterns from one-dimensional ^{1}H NMR spectra of small molecules, without relying on prior knowledge of the underlying molecular structure. This research was conducted within a collaborative project between the ZHAW School of Engineering and Bruker AG, supported by funding from Innosuisse, the Swiss Innovation Agency. We formulate the multiplet classification task as a one-dimensional segmentation problem defined over the full spectral axis, and develop a deep neural network that assigns to each spectral point a label corresponding to its splitting pattern class [52]. The architecture is composed of two main blocks. In the first, convolutional layers extract multi-scale features that capture local spectral characteristics. In the second, these features are integrated along the frequency axis through dense and long short-term memory (LSTM) layers, enabling the modeling of long-range dependencies and contextual correlations [53, 54, 55, 56, 57]. To train the proposed model, we introduce a pragmatic strategy for the generation of realistic synthetic spectral samples. Instead of performing computationally demanding quantum-mechanical simulations of spin-system Hamiltonians, spectra are constructed through random sampling of relevant spectral features, including peak positions, linewidths, intensities, and coupling patterns. This approach substantially reduces computational cost while retaining sufficient variability and realism for effective training in the context of splitting-pattern recognition.

When a resonance arises from a nucleus, or from a set of equivalent nuclei, coupled to multiple non-equivalent partners, the associated splitting tree becomes increasingly complex. Depending on the relative magnitudes of the coupling constants, certain components may be partially resolved or effectively indistinguishable, complicating the recognition of the underlying pattern. This variability must be explicitly accounted for in the construction of synthetic training data for deep learning models. To address this issue, we introduce, for each splitting class considered in our framework, a corresponding *multiplet phenotype*. This concept constrains the admissible combinations of couplings and rela-

tive intensities during signal generation, thereby imposing regularization on the synthetic data. The resulting procedure promotes intra-class consistency while preserving clear inter-class boundaries, facilitating stable and discriminative learning.

A further essential requirement in scientific research and development is the association of predictive outputs with a quantitative measure of confidence [58, 59]. Such uncertainty estimates enhance the trustworthiness of artificial intelligence–based classifications by indicating which predictions can be regarded as reliable and which require additional validation. In the present context, resonances exhibiting high predictive uncertainty may be assigned to a general multiplet class, which encompasses overlapping signals and higher-order patterns that cannot be unambiguously resolved within the adopted taxonomy. Moreover, uncertainty quantification provides insight into the learning process itself. Highly uncertain examples identify regions of the data space that are difficult for the model, thereby guiding the refinement of the training distribution and the representation of spectral features. Developed in close collaboration with the team at Bruker Switzerland AG, our framework evolved through continuous confrontation with application-driven requirements. The industrial partnership actively informed methodological choices, ensuring that the algorithm met practical expectations in terms of robustness to varying experimental conditions, increasing sample complexity, low signal-to-noise ratios, and common spectral distortions or artefacts.

The demand for automation becomes even more compelling in high-throughput application domains. Fragment-based drug discovery with ^{1}H NMR constitutes a paradigmatic example, in which large libraries of small molecules are screened for weak and strong binding interactions with a target protein, and where rapid, reproducible, automated and quantitatively grounded analysis becomes a methodological requirement. The ^{1}H NMR techniques most commonly employed in fragment screening include saturation transfer difference (STD) experiments [60], waterLOGSY [61], and $T_{1\rho}$-based methods [62]. More recently, the PEARLScreen experiments [13], based on transverse relaxation effects, has been introduced to enhance contrast in binding identification.

Low intrinsic sensitivity remains a fundamental limitation of NMR spectroscopy, often requiring high sample concentrations and extended acquisition times, a constraint that becomes particularly critical in the context of fragment screening. To overcome this limitation, several hyperpolarization strategies have been developed, including dynamic nuclear polarization (DNP) [63], parahydrogen-induced polarization (PHIP) [64], signal amplification by reversible exchange (SABRE) [65], and photochemically induced dynamic nuclear polarization (photo-CIDNP) [66, 67]. While exhibiting lower signal enhancements compared to other hyperpolarization methods, photo-CIDNP NMR offers several advantageous features. Its operational complexity is comparatively low, as polarization is triggered simply by illuminating the sample within the NMR probe. Moreover, the experiment can be performed in aqueous solution at room temperature, without chemical modification of the protein, and is compatible with low micromolar concentrations of both ligand and protein (down to 5 μM of ligand and 2 μM of target). The experiment requires only a few seconds, as sufficient signal-to-noise ratio can often be achieved within a single scan. Upon binding to a protein, the hyperpolarised signals are selectively quenched, enabling the identification of ligand–protein interactions and the estimation of binding affinities [12]. Notably, the quenching effect is proton-specific, providing site-resolved information that can be exploited to infer the binding epitope of the ligand.

NexMR AG, a spin-off from ETH Zürich, is a leading industrial actor in the application

of photo-CIDNP technology to drug discovery. Within an Innosuisse-funded collaborative project, we developed an automated analysis framework tailored to ^{1}H photo-CIDNP NMR spectra for the detection of ligand–protein binding and the estimation of binding strength. This task introduces additional complexity compared to the work described previously. Whereas the earlier part of this thesis focused on extracting intramolecular information from individual spectra, fragment screening requires the identification of intermolecular interactions. Methodologically, this shift is reflected in the need to correlate information across multiple spectra rather than analyzing a single experiment in isolation. The standard experimental setup comprises four spectra: a conventional ^{1}H NMR spectrum and a photo-CIDNP spectrum of the ligand alone, together with the corresponding pair acquired in the presence of the protein. Additional transverse relaxation experiments may further complement this dataset.

The proposed algorithm combines rule-based procedures with deep learning components to identify and localize hyperpolarised signals and to compare their amplitudes between ligand-only and ligand–protein spectra. From this comparison, a quenching measure is derived, proportional to the binding strength, such that larger signal attenuation corresponds to stronger interaction. Photo-CIDNP spectra exhibit distinctive features. Hyperpolarised signals can appear with both positive and negative (emissive) amplitudes, and strong resonances originating from auxiliary compounds required for sample preparation may introduce substantial baseline distortions. These distortions complicate global phase correction across the full spectral range, even under manual supervision. To mitigate these issues, we propose the use of the magnitude spectrum in place of the conventional real (absorptive) component. Although magnitude-mode processing is uncommon in high-resolution NMR spectroscopy, magnitude images are routinely employed in MRI [68]. We provide a systematic analysis of the modifications introduced by magnitude computation relative to the real spectrum and establish a dedicated preprocessing protocol aimed at obtaining symmetric peaks in magnitude mode, addressing previously reported asymmetry effects [69]. Within this framework, MuSe Net is employed to aggregate the automatically detected hyperpolarised peaks and to filter them according to their predicted multiplet class and associated uncertainty, thereby enabling reliable downstream quenching analysis despite the non-Gaussian noise characteristics of magnitude spectra. Hit prediction is performed without requiring prior knowledge of the molecular structure of the ligand. However, to map proton-specific quenching values onto the corresponding atomic sites, we developed a Bayesian assignment algorithm [70] adapted to the constraints of the present screening framework. Starting from the two-dimensional molecular structure of the ligand, chemical shifts are predicted using a graph neural network [71]. These predicted shifts are then probabilistically matched to the experimentally observed resonances, accounting for the fact that not all protons necessarily give rise to hyperpolarised signals and that certain detected peaks may originate from artefacts or unrelated species. The complete pipeline was implemented and evaluated on real experimental photo-CIDNP screening datasets comprising compounds from the partner company's fragment library. The method is computationally efficient and compatible with the high-throughput nature of photo-CIDNP screening, being able to automatically process approximately 2500 spectra in less than ten minutes while retrieving the relevant binding candidates. This performance demonstrates the framework's robustness and its immediate utility within operational screening environments.

The present thesis is organized as follows. **Chapter 2** provides a systematic descrip-

tion of the characteristic features of one–dimensional ^{1}H NMR spectra, with particular emphasis on those observables that enable the extraction of structural information at the molecular level. In addition, it presents the essential elements of the quantum–mechanical formalism underlying liquid–state ^{1}H NMR experiments, restricted to the concepts and approximations required for the subsequent methodological developments. **Chapter 3** outlines the principal stages of NMR data analysis that are most frequently required in practical applications, spanning the workflow from signal acquisition and spectral processing to spin–system identification and, ultimately, molecular structure elucidation. Both established automated methodologies and recent data–driven approaches based on deep learning are critically reviewed, with explicit attention to their underlying assumptions, domains of applicability, and inherent limitations. **Chapters 4 and 5** are devoted to the classification of resonances according to their splitting patterns. **Chapter 4** formulates multiplet classification as a segmentation problem and presents the implementation of a prototype convolutional neural network designed to address this task. **Chapter 5** introduces the probabilistic deep learning framework *MuSe Net* and provides a comprehensive evaluation of its performance, including a detailed analysis of predictive accuracy, calibration behaviour, robustness across heterogeneous spectral conditions, and its uncertainty quantification properties. **Chapter 6** outlines the theoretical foundations of nuclear spin hyperpolarization, with particular emphasis on the photo–CIDNP effect in NMR spectroscopy, and discusses its application to fragment–based screening experiments. **Chapter 7** presents the automated algorithm developed for ligand–protein binding detection and reports its performance on real experimental datasets. The chapter further details how the method can be integrated into the annotation workflow, specifying the operational steps, decision criteria, and practical constraints that govern its effective deployment. Finally, **Chapter 8** discusses the contributions of this work and situates the obtained results within the landscape of existing methodologies and scientific applications, while outlining open challenges and concrete directions for future research.

Chapter 2

Observing the spectra

A one-dimensional NMR spectrum is a representation in which the detected signal intensity, corresponding to net absorption or emission, is plotted along the vertical axis as a function of frequency on the horizontal axis. The underlying signal arises from nuclear magnetization evolving in an external magnetic field and is detected as an induced electromagnetic signal. Spectral features appear at specific resonance frequencies determined by the interaction between nuclear magnetic moments and the field. The nuclear magnetic moment μ is proportional to the nuclear spin angular momentum $\hat{\mathbf{I}}$ through the gyromagnetic ratio γ, such that

$$\boldsymbol{\mu} = \gamma \hat{\mathbf{I}}. \tag{2.1}$$

For nuclei with a positive gyromagnetic ratio ($\gamma > 0$), the magnetic moment and the spin angular momentum are therefore oriented in the same direction (*spin polarization* [72]). The presence of a non-zero nuclear spin, and hence of a magnetic moment, is the fundamental property that enables NMR detection. Accordingly, in the context of NMR spectroscopy, the term *spin* is customarily used as a shorthand to denote the nuclei of the molecule under investigation that possess a non-zero spin quantum number.

Nuclear spin angular momenta are randomly oriented. In the presence of large external magnetic field, for example when placing the sample inside the NMR magnet, the spin angular momentum vectors execute a rapid, nearly constant-angle precessional motion about the field direction with a characteristic angular frequency called the Larmor frequency:

$$\omega_0 = -\gamma B_0. \tag{2.2}$$

However, weak and fluctuating magnetic fields originating from the thermal environment progressively perturb this idealized motion. The fast precession, occurring on a timescale set by ω_0^{-1} and typically of order nanoseconds, is thus accompanied by a much slower stochastic reorientation of the spins driven by molecular motions, a process that may extend over timescales of seconds. This slow thermal dynamics is weakly biased toward spin orientations in which the magnetic moment is aligned parallel to the external field. As a consequence, the spin system relaxes toward a stationary, anisotropic distribution of populations known as thermal equilibrium, giving rise to a small net magnetization along the field direction, termed longitudinal magnetization. The approach to this equilibrium state is characterized by the time constant T_1, referred to as the longitudinal or spin–lattice relaxation time.

In the simplest NMR experiment [72], a radiofrequency (r.f.) pulse generated by the r.f. synthesizer and delivered to the sample through the r.f. coil perturbs the thermal

equilibrium of the spin system. For example, a $\pi/2$ pulse applied about the x axis rotates the equilibrium spin polarization into the transverse plane, producing a magnetization oriented along the $-y$ direction. When the r.f. field is switched off, this transverse magnetization undergoes free precession in the xy plane at the Larmor frequency:

$$\begin{aligned} M_y &= -M_{eq} cos(\omega^0 t) exp\{-t/T_2\}, \\ M_x &= M_{eq} sin(\omega^0 t) exp\{-t/T_2\}. \end{aligned} \tag{2.3}$$

The resulting time-dependent magnetic flux induces the NMR signal in the detection coil of the probe, which is transmitted to the quadrature receiver and recorded as a complex-valued signal.

The free induction decay (FID) denotes the NMR signal acquired in the time domain after application of a radiofrequency pulse. It is conveniently described as a linear superposition of elementary signal contributions,

$$s(t) = \sum_l s_l(t), \qquad s_l(t) = a_l \exp[(\mathrm{i}\,\omega_l - \lambda_l)t]\,, \tag{2.4}$$

where each term $s_l(t)$ corresponds to a single spectral component. In the most general case, the components differ in their angular frequencies ω_l, their exponential decay constants λ_l, and their complex amplitudes a_l. This representation is valid for the NMR signal generated by an arbitrary pulse sequence. The resonance frequencies ω_l and relaxation rates λ_l are determined exclusively by the spin dynamics during signal acquisition, whereas the amplitudes a_l encode the state of the spin system at the onset of detection. Writing the complex amplitude in polar form,

$$a_l = |a_l| \exp(\mathrm{i}\,\phi_l)\,, \tag{2.5}$$

the modulus $|a_l|$ defines the intensity of the corresponding signal component, while the argument ϕ_l specifies its phase.

The transformation from the time domain to the frequency domain is achieved by application of Fourier transform [73]. The Fourier transform decomposes the composite time-domain signal into its individual frequency components, thereby rendering the spectral content in a form that is directly interpretable as resonance lines at well-defined frequencies. Crucially, this mathematical operation does not increase the theoretical information content of the signal, but merely reorganizes the information already present in the FID into a representation that is more accessible for chemical interpretation and quantitative analysis.

The mathematical definition of the Fourier Transform is the following

$$S(\omega) = \int_0^\infty s(t) \exp\{-i\omega t\}\, dt, \tag{2.6}$$

Substituting the decomposition of the FID given in Eq. 2.4 into this definition yields

$$S(\omega) = \sum_l S_l(\omega), \qquad S_l(\omega) = a_l \cdot \frac{1}{\lambda_l + i(\omega - \omega_l)} \equiv a_l\, \mathcal{L}(\omega; \omega_l, \lambda_l) \tag{2.7}$$

where ω_l is the resonance frequency of the l-th peak and λ_l is its decay rate, which governs

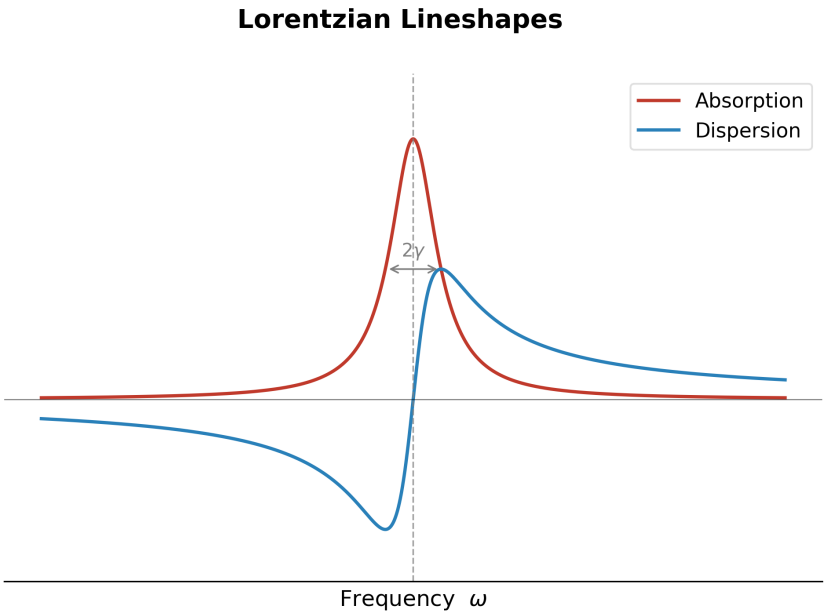


Figure 2.1: Real (absorptive, top) and imaginary (dispersive, bottom) parts of the complex Lorentzian lineshape as a function of angular frequency ω. The resonance is centered at ω_0, indicated by the vertical line.

the linewidth. Each peak has the lineshape of the *complex Lorentzian* (see Fig. 2.1)

$$\begin{aligned} \mathcal{L}(\omega;\omega_l,\lambda_l) &= \mathcal{A}(\omega;\omega_l,\lambda_l) + \mathrm{i}\mathcal{D}(\omega;\omega_l,\lambda_l) \\ \mathcal{A}(\omega;\omega_l,\lambda_l) &= \mathrm{Re}\{\mathcal{L}(\omega;\omega_l,\lambda_l)\} = \frac{\lambda_l}{\lambda_l^2+(\omega-\omega_l)^2} \\ \mathcal{D}(\omega;\omega_l,\lambda_l) &= \mathrm{Im}\{\mathcal{L}(\omega;\omega_l,\lambda_l)\} = -\frac{\omega-\omega_l}{\lambda_l^2+(\omega-\omega_l)^2} \end{aligned} \tag{2.8}$$

The real part $\mathcal{A}(\omega;\omega_l,\lambda_l)$ corresponds to the absorptive Lorentzian component, centered at ω_l with half-width at half-maximum equal to λ_l, while the imaginary part $\mathcal{D}(\omega;\omega_l,\lambda_l)$ represents the dispersive contribution.

When the amplitudes a_l of all signal components are purely real, the real part of the spectrum, $\Re\{S(\omega)\}$, contains only absorptive Lorentzian lineshapes, while the imaginary part, $\Im\{S(\omega)\}$ consists exclusively of dispersive Lorentzians. Since absorptive lines provide superior spectral resolution, NMR spectra are typically displayed using the real component alone. In the general case, however, the amplitudes a_l are complex. Both the real and imaginary parts of the spectrum then become linear combinations of absorptive and dispersive contributions. This mixing distorts the ideal Lorentzian profile and leads to asymmetric peak shapes; the spectrum is said to be out of phase. Several procedures have been developed to correct such phase distortions, and these methods will be discussed in the following chapter.

2.1 Resonances characterization

The signal features in the spectrum appear either as isolated peaks, commonly referred to as singlets, or they are split into clusters of closely spaced peaks, commonly referred to as multiplets. Each resonance in the spectrum originates from a magnetic nucleus

or from a functional group within the molecule [21]. The characteristic parameters of these resonances, namely their chemical shift, splitting pattern, and integrated intensity, are determined by the local molecular environment and the connectivity between nuclei. Taken together, these spectral features provide a reproducible and molecule-specific signature, allowing one-dimensional NMR spectra to serve as a structural fingerprint of the compound under investigation.

The absolute frequency at which an NMR signal is observed depends on both the nuclear isotope under investigation and the strength of the applied magnetic field. Different nuclei (e.g. ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P etc.) possess distinct magnetic properties, resulting in characteristic resonance frequencies even under identical field conditions. In addition, for a given nucleus, the resonance frequency scales linearly with the applied magnetic field strength, such that changes in field intensity lead to proportional changes in observed frequencies. As a consequence, absolute resonance frequencies are not intrinsic spectral coordinates, motivating the use of normalized representations when comparing spectra acquired under different experimental conditions.

The position of a spectral peak is specified by measuring its frequency offset relative to a reference signal and normalizing this offset by the reference frequency. Because this definition involves the ratio of two frequencies, the dependence on the applied magnetic field cancels. The resulting quantity is called the chemical shift, defined as:

$$\delta = \frac{(\nu - \nu_{ref})}{\nu_{ref}} \times 10^6 \text{ ppm}, \tag{2.9}$$

where ν is the absolute frequency of the observed peak, ν_{ref} is the absolute frequency of the reference signal, and the chemical shift δ is expressed in parts per million (ppm). The most commonly adopted reference for chemical shift measurements is tetramethylsilane (TMS). In one-dimensional ^{1}H NMR spectra, the TMS singlet is conventionally placed at the origin of the frequency axis. Under this convention, signals located to the right of the reference are assigned negative chemical shift values, while those located to the left are assigned positive values. From a physical viewpoint, the chemical shift reflects the extent to which the local magnetic field experienced by a nucleus differs from the externally applied field as a result of its electronic environment. In ^{1}H NMR spectroscopy, this modification arises from the response of surrounding electrons to the applied magnetic field, which induces secondary magnetic fields at the position of the nucleus. When the electronic environment reduces the effective magnetic field at the nucleus, the resonance frequency decreases, resulting in a smaller chemical shift value. This effect is referred to as shielding. Conversely, when the electronic environment enhances the effective magnetic field at the nucleus, the resonance frequency increases, leading to a larger chemical shift value; this is termed deshielding. Because the electronic environment depends on the local molecular structure, these shielding and deshielding effects are not arbitrary but follow systematic trends across different chemical environments. As a consequence, protons associated with different functional groups tend to resonate in distinct regions of the chemical shift scale. The positions of peaks in a one-dimensional ^{1}H NMR spectrum therefore provide qualitative information about molecular structure, allowing spectral features to be associated with broad classes of chemical environments.

As introduced above, spectral features may appear either as isolated peaks or as groups of peaks: isolated peaks arise from nuclei that do not exhibit observable coupling to other magnetic nuclei, whereas multiplets originate from nuclei that are coupled to one or more neighboring magnetic nuclei (see Figure 2.2).

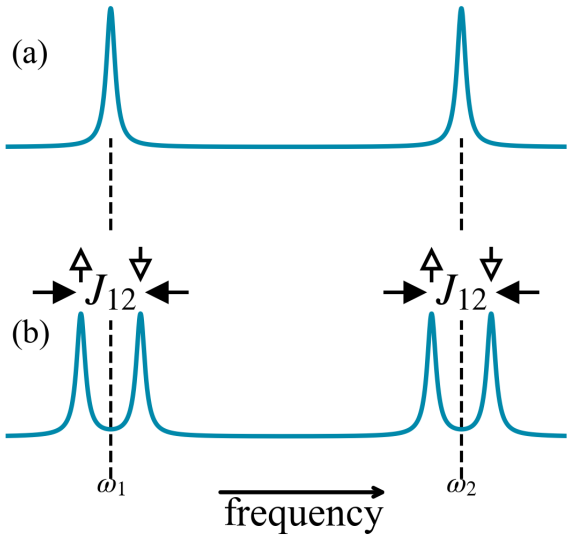


Figure 2.2: Effect of scalar coupling on NMR signals. (a) Two uncoupled nuclei produce single resonance peaks at frequencies ω_1 and ω_2. (b) Scalar coupling (J_{12}) splits each resonance into a doublet separated by the coupling constant.

Scalar, or J, coupling between nuclear spins is mediated through chemical bonds and therefore provides information on the relative proximity and connectivity of nuclei within a molecular framework. The strength of this interaction is independent of the applied magnetic field and is conventionally expressed in hertz. Because scalar coupling is transmitted through chemical bonds, its magnitude generally decreases as the number of intervening bonds increases; accordingly, a superscript notation is used to indicate the coupling order, with ${}^{n}J$ denoting a coupling mediated across n chemical bonds.

Within a limited number of chemical bonds, hydrogen nuclei can exhibit observable scalar coupling to one another. In routine ^{1}H NMR spectroscopy, such couplings are therefore dominated by proton–proton interactions. Although protons may in principle couple to carbon nuclei, the most abundant carbon isotope, ^{12}C, has zero nuclear spin and does not give rise to coupling splittings. As a result, proton–carbon coupling is generally not observed in standard ^{1}H spectra, except for weak satellite peaks arising from the low natural abundance of the spin-active isotope ^{13}C. These carbon satellites appear as pairs of weak lines symmetrically displaced about the central resonance and exhibit the same splitting pattern as the main proton signal, reflecting coupling to the low-abundance ^{13}C isotope. Scalar couplings between hydrogen nuclei and isotopes with spin greater than $\frac{1}{2}$ are usually not observed. These nuclei interact with electric field gradients that are modulated by molecular rotation, causing large magnetic moment fluctuations and very fast spin–lattice relaxation, which effectively averages out the corresponding J-couplings.

The nuclear isotopes most frequently targeted in liquid-state NMR experiments, including ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P, possess a nuclear spin quantum number of $\frac{1}{2}$. Therefore, in the remainder of this thesis, only the case of spin-$\frac{1}{2}$ nuclei will be considered. When two spin-$\frac{1}{2}$ nuclei are coupled, the resonance associated with each nucleus is split

symmetrically about its chemical shift position into two components, giving rise to a doublet.

If a nucleus is coupled to more than one neighboring spin, the observed resonance undergoes successive splittings, and the resulting multiplet pattern reflects the combined effect of the individual couplings (see Figure 2.3). These patterns can be constructed conceptually by applying each coupling interaction in turn, with each additional coupled nucleus introducing a further level of splitting. For example, coupling of a spin-$\frac{1}{2}$ nucleus to two distinct spin-$\frac{1}{2}$ neighbors, with coupling constants J_1 and J_2, produces a four-line multiplet, commonly referred to as a doublet of doublets, in which each pair of lines is separated by its respective coupling constant. More complex multiplets arise analogously when additional couplings are present, with their overall structure encoding information about the local bonding environment of the observed nucleus.

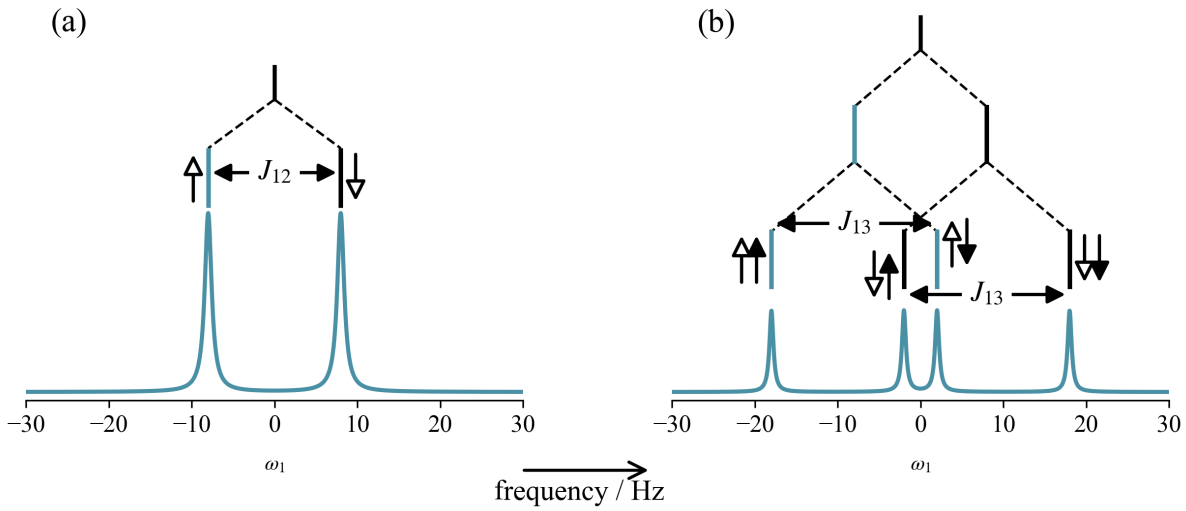


Figure 2.3: Splitting patterns arising from scalar coupling in NMR spectroscopy. (a) A two-spin system produces a doublet with splitting J_{12}. (b) In a three-spin system, additional couplings (J_{13}) lead to further splitting, resulting in a more complex multiplet pattern. The splitting tree illustrates how individual spin states contribute to the observed spectral lines.

A signal originated by a nuclei coupled to n spin-$\frac{1}{2}$ nuclei gives rise, in general, to a multiplet consisting of up to 2^n spectral lines symmetrically distributed about the central frequency δ. In practice, the number of resolved lines may be reduced due to degeneracies that arise when two or more coupling constants are equal, or when specific relationships exist among them. Nevertheless, the total number of underlying intensity contributions always remains equal to 2^n, regardless of whether individual lines are resolved or superimposed.

This can be illustrated by returning to the case of a resonance coupled to $n = 2$ spin-$\frac{1}{2}$ nuclei. When the two coupling constants are distinct, the signal is split into four lines of equal intensity, forming a doublet of doublets. If the two coupling constants are identical, the two central lines coincide, resulting in a three-peak multiplet with relative intensities 1:2:1, commonly referred to as a triplet. Although the number of observable peaks is reduced from four to three, the total number of components remains four, reflecting the underlying coupling to two spin-$\frac{1}{2}$ nuclei.

Each of the 2^n intensity components that constitute a first-order multiplet corresponds to a distinct frequency offset relative to the central chemical shift δ, obtained by combining the individual coupling constants with appropriate signs. In particular, the most downfield and upfield components arise when all coupling contributions add constructively or destructively, respectively, such that the outermost line positions are shifted by $\pm\frac{1}{2}(J_1 + J_2 + \cdots + J_n)$ from δ. As a consequence, the total frequency span of the multiplet is equal to the sum of the individual coupling constants.

For generality, any multiplet resulting from coupling to n spin-$\frac{1}{2}$ nuclei may be regarded as arising from the successive splitting of a single resonance into n doublets, independent of whether the coupled nuclei are chemically or magnetically equivalent. When equivalence is present, degeneracies occur and the resulting patterns are conventionally described using simplified nomenclature, such as triplets, quartets, or higher-order multiplets, corresponding to coincident splittings with identical coupling constants (see Figure 2.4).

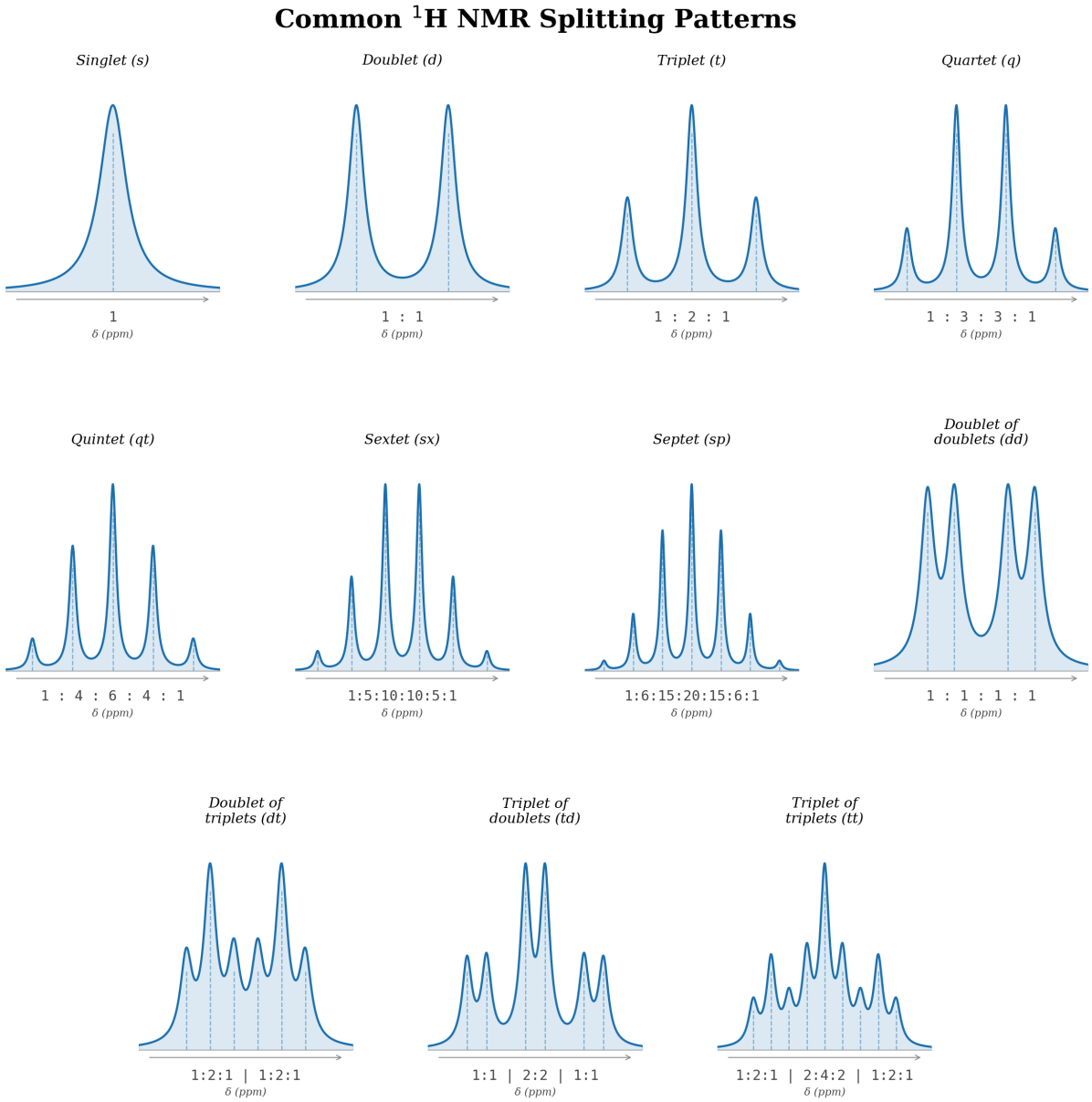


Figure 2.4: Common ^{1}H NMR splitting patterns arising from scalar (spin–spin) coupling. The multiplicities shown include singlet (s), doublet (d), triplet (t), quartet (q), quintet (qt), sextet (sx), septet (sp), and higher-order patterns such as doublet of doublets (dd), doublet of triplets (dt), triplet of doublets (td), and triplet of triplets (tt). The relative peak intensities follow Pascal's triangle rule.

In addition to chemical shift, splitting pattern, and coupling constants, the peak integral constitutes another important descriptor of NMR resonances. The integral of a resonance or multiplet is proportional to the number of nuclei contributing to the signal and is independent of both the splitting pattern and the linewidth. Linewidth is commonly quantified as the full width at half maximum (FWHM) of the peak and reflects the frequency range over which the signal intensity is distributed. For a fixed integral, an increase in the number of split components or an increase in linewidth leads to a redistribution of intensity over a larger number of frequency points, resulting in reduced peak heights.

Structural symmetry within a molecule reduces both the number of observable res-

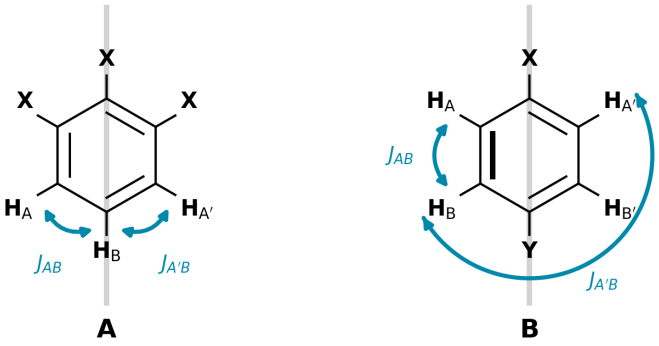


Figure 2.5: Comparison between chemical and magnetic equivalence. In panel A, the mirror plane (grey line) renders nuclei H_A and H'_A chemically equivalent, with identical chemical shifts; equality of the couplings J_{AB} and $J_{A'B}$ further implies magnetic equivalence. In panel B, H_A and H'_A (as well as H_B and H'_B) remain chemically equivalent, but the unequal couplings $J_{AB} \neq J_{A'B}$ break magnetic equivalence.

onances and the complexity of their splitting patterns. Atoms, or identical functional groups, that experience the same chemical environment are said to be chemically equivalent (see Figure 2.5). This verifies when the spins are of the same isotopic species and it exists a molecular symmetry operation that exchanges the two spin. When the spins are chemically equivalent and they also share identical coupling relationships with all other magnetic nuclei in the molecule, they are said to be magnetically equivalent. Chemically equivalent nuclei or functional groups share the same chemical shift. However, the converse is not generally true. Spins that share the same chemical shift are not necessarily chemically equivalent: identical resonance frequencies may arise accidentally and need not reflect any underlying molecular symmetry. Finally, the couplings between magnetically equivalent nuclei are not observed in the NMR experiment.

2.2 First- and Second-Order ^{1}H NMR Spectra

On the basis of chemical and magnetic equivalence, ^{1}H NMR spectra are commonly divided into first-order and second-order regimes. First-order spectral behavior is observed when two conditions are simultaneously met. The first is that the frequency separation between coupled resonances is sufficiently large compared to the corresponding scalar coupling constants, such that $\Delta\nu/J \gg 1$ when expressed in consistent frequency units; this condition defines the weak-coupling regime. The second is that the nuclei involved in each coupling group are magnetically equivalent. Under these conditions, multiplet structures follow the empirical $n+1$ rule, with relative peak intensities well approximated by binomial coefficients (e.g., 1:1 for a doublet, 1:2:1 for a triplet, and 1:3:3:1 for a quartet). Moreover, both the chemical shift and the coupling constant can be determined directly from the spectrum, with the chemical shift given by the center of the multiplet and the coupling constant by the spacing between adjacent lines.

When one or both of these conditions are not fulfilled, the system departs from the weak-coupling regime and enters the strong-coupling regime, in which the frequency separation between coupled resonances becomes comparable to the coupling constant. Spec-

tra in this regime are classified as second-order and exhibit departures from first-order behavior, including an increased number of lines, non-binomial and asymmetric intensity distributions, and shifts in line positions. As a result, chemical shifts and coupling constants can no longer be extracted by straightforward inspection, and spectral interpretation becomes correspondingly more involved. In practice, spectra acquired at high magnetic field strengths are more likely to satisfy the weak-coupling condition, whereas low-field measurements, such as those performed on benchtop spectrometers, which are generally more affordable, more frequently exhibit strong-coupling effects due to reduced frequency dispersion.

2.3 Pople Nomenclature of Spin Systems

The set of mutually coupled nuclear spins within a single molecule is referred to as a *molecular spin system*. Such systems are conventionally denoted using letters of the alphabet, chosen to reflect relative chemical shift separations; this convention is formally known as Pople nomenclature [74]. Spins with widely separated chemical shifts are assigned letters that are distant in the alphabet, a convention that typically corresponds to weak coupling. Conversely, spins with similar chemical shifts are denoted by adjacent letters; for example, the notation AB indicates a strongly coupled two-spin system.

Magnetically equivalent spins are indicated by a numerical subscript, whereas spins that are chemically equivalent but magnetically inequivalent are distinguished by primes. As an illustration, the six protons in benzene are magnetically equivalent and are therefore denoted A_6. In ethyl chloride, the five protons form two groups of magnetically equivalent spins, comprising three and two protons, respectively. Since their chemical shift difference is usually sufficient to ensure weak coupling, the corresponding spin system is denoted A_3X_2. A common example of chemically equivalent but magnetically inequivalent spins is provided by the $AA'XX'$ spin system of a $X{-}CH_2{-}CH_2{-}Y$ fragment, where each methylene group contains two nonequivalent protons coupled to each other and to the neighboring group.

2.4 Quantum Mechanical Description of 1D ^{1}H NMR Spectra

In a fully general formulation [72], the physical state of a macroscopic sample investigated in a ^{1}H NMR experiment may be represented by a single many-body wavefunction $|\psi_{\text{full}}(t)\rangle$. This state vector contains, in principle, complete information about the system, including the spatial coordinates and momenta of all nuclei and electrons, as well as their associated spin degrees of freedom. Its time evolution is governed by the time-dependent Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}|\psi_{\text{full}}(t)\rangle = \hat{\mathcal{H}}_{\text{full}}|\psi_{\text{full}}(t)\rangle, \tag{2.10}$$

where $\hat{\mathcal{H}}_{\text{full}}$ denotes the exact Hamiltonian of the interacting many-particle system. Although formally complete, this equation is of no practical use for NMR, as it cannot be solved for any realistic system.

A pronounced separation of timescales exists between electronic and nuclear degrees of freedom. Electronic motions evolve on timescales that are several orders of magnitude

shorter than those associated with nuclear spin dynamics. As a consequence, nuclear spins effectively experience only time-averaged magnetic and electric fields generated by the electrons, and the back-action of the nuclear spins on the electronic degrees of freedom can be neglected.

This separation of timescales permits a substantial reduction of the quantum-mechanical description, in which only the nuclear spin degrees of freedom are retained explicitly. The resulting approximate dynamics are described by

$$\frac{d}{dt}|\psi_{\text{spin}}\rangle(t) = -\frac{i}{\hbar}\,\hat{\mathcal{H}}_{\text{spin}}\,|\psi_{\text{spin}}\rangle(t), \tag{2.11}$$

where $|\psi_{\text{spin}}\rangle$ denotes the quantum state of the nuclear spins and $\hat{\mathcal{H}}_{\text{spin}}$ is the nuclear spin Hamiltonian. This Hamiltonian contains only terms that depend on the orientations of the nuclear spin angular momenta, while the influence of electronic and molecular degrees of freedom enters implicitly through effective, time-averaged interactions.

The assumption underlying this reduction is commonly referred to as the *spin Hamiltonian hypothesis*. Under ordinary experimental conditions, the hierarchy of timescales on which it rests is well satisfied, rendering the nuclear spin Hamiltonian a robust and widely applicable framework for the theoretical description of NMR.

In the following, the Hamiltonian operator $\hat{\mathcal{H}}$ is understood to denote the *nuclear spin Hamiltonian*, and the quantum states $|\psi\rangle$ are taken to represent nuclear spin states only.

For a nucleus j in an external electromagnetic environment, the interaction Hamiltonian $\hat{\mathcal{H}}_j$ can, in principle, be decomposed into electric and magnetic contributions,

$$\hat{\mathcal{H}}_j = \hat{\mathcal{H}}_j^{\text{elec}} + \hat{\mathcal{H}}_j^{\text{mag}}, \tag{2.12}$$

reflecting the presence of an electric charge and a magnetic dipole moment associated with the nuclear spin.

For nuclei with spin quantum number $I = \frac{1}{2}$, all electric multipole moments beyond the monopole vanish. The monopole term corresponds to the total nuclear charge and does not give rise to any orientation-dependent interaction with external electric fields; as such, it has no direct influence on NMR observables. Consequently, for spin-$\frac{1}{2}$ nuclei the electric contribution $\hat{\mathcal{H}}_j^{\text{elec}}$ can be neglected, and the nuclear spin Hamiltonian is entirely determined by magnetic interactions.

In contrast, nuclei with spin $I > \frac{1}{2}$ possess a non-vanishing electric quadrupole moment, which couples to the local electric field gradient at the nuclear site. This quadrupolar interaction constitutes an additional term in the nuclear spin Hamiltonian and strongly influences the corresponding NMR spectra. Nuclei with $I > \frac{1}{2}$ are therefore commonly referred to as *quadrupolar nuclei*.

The interactions experienced by a nucleus may be classified as either external, arising from the experimental apparatus, or internal, originating from the surrounding molecular environment. For spin-$\frac{1}{2}$ nuclei, both classes of interactions are magnetic in nature. A characteristic and experimentally crucial feature of NMR is that the externally applied magnetic interactions are typically much stronger than the internal interactions. As a consequence, nuclear spins couple more strongly to the fields generated by the spectrometer than to those arising from their local molecular environment. This hierarchy of interactions is what governs both the sensitivity and the interpretability of NMR spectra.

Modern NMR spectrometers can supply several types of external magnetic field. The dominant contribution is a strong, highly homogeneous static field B_0, generated by

a superconducting solenoid. In addition, an oscillating radiofrequency field $\mathbf{B}_{\mathrm{RF}}(t)$ is produced by the radiofrequency coil in the probe; under standard operating conditions, this field is designed to be spatially homogeneous over the sample volume. Spectrometers equipped with gradient hardware may further generate a magnetic field $\mathbf{B}_{\mathrm{grad}}(\mathbf{r}, t)$, which is much weaker than B_0, depends explicitly on the spatial coordinate $\mathbf{r}$, and may be varied in a controlled manner in time. Such gradient fields are produced by dedicated gradient coils and are essential for spatial encoding and coherence selection.

The internal nuclear spin Hamiltonian arises from interactions within the sample and, in diamagnetic systems, contains several distinct contributions. Chemical shift interactions describe the indirect coupling of the nuclear spins to the external magnetic field mediated by the surrounding electrons. Quadrupolar couplings account for the interaction of nuclei with spin $I > \frac{1}{2}$ and a non-vanishing electric quadrupole moment with local electric field gradients. Direct dipole–dipole couplings represent the through-space magnetic interaction between nuclear magnetic moments. Scalar, or J-couplings describe indirect magnetic interactions between nuclear spins transmitted via the electronic structure. Finally, spin–rotation interactions arise from magnetic fields generated by rotational motion of the molecules and couple to the nuclear magnetic moments. Together, these internal interactions determine the fine structure of NMR spectra and encode detailed information about molecular structure and dynamics.

The internal Hamiltonian can be simplified by means of two standard procedures. The first is the *secular approximation*, which allows the internal Hamiltonian to be replaced by a simplified form that is block diagonal. This approximation exploits the fact that the total Hamiltonian may be written as the sum of an external and an internal contribution,

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{\mathrm{ext}} + \hat{\mathcal{H}}_{\mathrm{int}}, \tag{2.13}$$

with the external interaction typically much stronger than the internal one. Both operators are assumed to be Hermitian.

One may therefore express $\hat{\mathcal{H}}_{\mathrm{int}}$ in the eigenbasis of $\hat{\mathcal{H}}_{\mathrm{ext}}$, defined by

$$\hat{\mathcal{H}}_{\mathrm{ext}}|n\rangle = a_n|n\rangle. \tag{2.14}$$

In this representation, the matrix elements of the internal Hamiltonian are $h_{mn} = \langle m|\hat{\mathcal{H}}_{\mathrm{int}}|n\rangle$. The secular approximation consists in neglecting those off-diagonal elements for which

$$|h_{mn}| \ll |a_m - a_n|, \tag{2.15}$$

since such terms generate rapidly oscillating contributions that average to zero on the timescale of the spin dynamics. As a result, only matrix elements connecting degenerate or nearly degenerate eigenstates of $\hat{\mathcal{H}}_{\mathrm{ext}}$ are retained, yielding a block diagonal effective Hamiltonian.

The second simplification procedure is *motional averaging*. When molecules undergo rapid motion, the internal interaction terms become explicitly time dependent and fluctuate on the timescale of the nuclear spin dynamics. If the molecular motion is sufficiently fast, the time-dependent internal Hamiltonian $\hat{\mathcal{H}}_{\mathrm{int}}(t)$ may be replaced by its time-averaged form, denoted $\overline{\hat{\mathcal{H}}_{\mathrm{int}}}$. In this approximation, those components of $\hat{\mathcal{H}}_{\mathrm{int}}(t)$ whose time average vanishes are discarded, as their contributions average to zero over the relevant timescale.

The use of a motionally averaged spin Hamiltonian is generally well justified in gases and liquids, where molecular reorientation is rapid, and becomes invalid only when molecular motion is sufficiently slow for the time dependence of the interactions to be resolved by the nuclear spins.

2.4.1 Spin Hamiltonian of 1D ^{1}H NMR

In the absence of an applied radiofrequency field, the spin dynamics of a set of weakly coupled nuclei in an isotropic liquid are governed by a Hamiltonian consisting of the static external magnetic field contribution $\hat{\mathcal{H}}^{\text{static}}$ together with the chemical shift $\hat{\mathcal{H}}^{\text{cs}}$ and scalar (J-)coupling interactions $\hat{\mathcal{H}}^{\text{J}}$:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}^{\text{static}} + \hat{\mathcal{H}}^{\text{cs}} + \hat{\mathcal{H}}^{\text{J}}. \tag{2.16}$$

In this expression, intermolecular dipole–dipole and spin rotation interactions do not contribute appreciably. Short-range interactions are averaged to zero by the rapid translational motion of the molecules, while long-range interactions, although not strictly eliminated by isotropic averaging, are typically several orders of magnitude weaker than the retained terms and can be neglected in most practical situations.

It is standard practice to introduce a coordinate system in which the static magnetic field is aligned with the z-axis, thereby defining the laboratory reference frame. In this frame, the interaction of a nuclear spin j with the static longitudinal field B_0 is described by the Hamiltonian

$$\hat{\mathcal{H}}_j^{\text{static}} = -\gamma_j B_0 \hat{I}_{jz}, \tag{2.17}$$

where γ_j is the gyromagnetic ratio of the nucleus, $\hat{I}_{jz}$ is its spin angular momentum, and $\omega_j^0 \equiv -\gamma_j B_0$ denotes its Larmor angular frequency. This term is known as the nuclear Zeeman interaction and constitutes the dominant contribution to the spin Hamiltonian in high-field NMR.

The chemical shift arises from an indirect interaction between the nuclear spins and the applied magnetic field mediated by the molecular electrons. The external field B_0 induces circulating currents in the electronic cloud, which in turn generate a secondary magnetic field $\mathbf{B}_{\text{induced}}$. As a result, nucleus j experiences a local field

$$\mathbf{B}_{\text{loc},j} = \mathbf{B}_0 + \mathbf{B}_{\text{induced},j}. \tag{2.18}$$

The induced field is typically of order $10^{-6} - 10^{-5} B_0$, sufficiently small compared to the applied field yet large enough to produce measurable shifts in the nuclear precession frequencies. In general, the chemical shift is a second-rank tensor represented by a 3×3 matrix $\boldsymbol{\delta}_j$. In isotropic liquids, rapid molecular tumbling averages out its anisotropic components, so that only the isotropic contribution is observed:

$$\mathbf{B}_{\text{induced},j} = \boldsymbol{\delta}_j \cdot \mathbf{B}_0 = \delta_{\text{iso},j} B_0. \tag{2.19}$$

The chemical shift correlates strongly with electronegativity: electronegative substituents such as O, F, or Cl withdraw electron density from neighbouring atoms, thereby increasing the local magnetic field at nearby nuclei and leading to larger values of δ. More generally, the chemical shift is primarily determined by the intramolecular electronic structure, but it also contains a non-negligible intermolecular contribution. As a result, the observed chemical shift depends to some extent on the surrounding environment, giving rise, for example, to small but measurable variations between different solvents. The dominant contribution to the chemical shift typically arises from low-lying electronic excited states. Since heavier atoms possess a denser manifold of such states than lighter ones, they exhibit a substantially wider chemical shift range. Accordingly, the chemical shift dispersion for ^{1}H is usually limited to approximately 10 ppm, whereas for ^{13}C it extends to about 200 ppm.

Scalar (J-)coupling constitutes a direct spectroscopic signature of chemical bonding. Two nuclear spins exhibit an observable J-coupling only when they are connected through a limited number of chemical bonds, including, in some cases, hydrogen bonds. In its most general form, the coupling between spins j and k is described by

$$\hat{\mathcal{H}}^{\text{full}}_{J,jk} = 2\pi \mathbf{J}_{jk} \hat{\mathbf{I}}_j \cdot \hat{\mathbf{I}}_k, \tag{2.20}$$

where $\mathbf{J}_{jk}$ is the J-coupling tensor, represented by a real 3×3 matrix, and $\hat{\mathbf{I}}$ is the nuclear spin angular momentum.

In an isotropic liquid, rapid molecular tumbling averages the tensorial interaction to its scalar component. Under the additional assumption of weak coupling, the interaction reduces to

$$\hat{\mathcal{H}}_{J,jk} = 2\pi J_{jk}\, \hat{I}_{jz} \hat{I}_{kz}. \tag{2.21}$$

Collecting the static Zeeman, chemical shift, and scalar coupling contributions, the spin Hamiltonian in the absence of radiofrequency fields may be written as

$$\hat{\mathcal{H}}^0 = \sum_j \omega^0_j\, \hat{I}_{jz} + \sum_{j<k} 2\pi J_{jk}\, \hat{I}_{jz} \hat{I}_{kz}, \tag{2.22}$$

where the sums run over all magnetically inequivalent spins that are not subject to rapid dynamical averaging. The isotropic chemical shift frequency is defined as

$$\omega^0_j = -\gamma_j B_0 \big(1 + \delta_{\text{iso},j}\big). \tag{2.23}$$

2.4.2 Energy states

An ensemble of weakly coupled spin-$\frac{1}{2}$ nuclei possesses 2^N stationary states, which are given by the direct products of the Zeeman eigenstates of the individual spins. These single-spin Zeeman eigenstates are defined as the eigenstates of the z-component of the nuclear angular momentum operator, $\hat{\mathbf{I}}_z$, which is taken to be aligned with the direction of the static magnetic field B_0.

For a spin-$\frac{1}{2}$ nucleus, the Zeeman eigenstates are

$$\begin{aligned} |\alpha\rangle &= |\frac{1}{2}, +\frac{1}{2}\rangle, \quad \rightarrow \quad \hat{\mathbf{I}}_z|\alpha\rangle = +\frac{1}{2}|\alpha\rangle, \\ |\beta\rangle &= |\frac{1}{2}, -\frac{1}{2}\rangle, \quad \rightarrow \quad \hat{\mathbf{I}}_z|\beta\rangle = -\frac{1}{2}|\beta\rangle, \end{aligned} \tag{2.24}$$

where the kets $|I, m\rangle$ denote simultaneous eigenstates of $\hat{\mathbf{I}}^2$ and $\hat{\mathbf{I}}_z$.

More generally, for a nucleus with spin quantum number I, the operator $\hat{\mathbf{I}}_z$ has $2I+1$ eigenvalues

$$m \in \{-I, -I+1, \ldots, I-1, I\}, \tag{2.25}$$

which label the allowed Zeeman states of the individual spin.

Accordingly, the states of an N-spin system may be written as product states of the single-spin Zeeman eigenstates,

$$|r\rangle = |m_1, m_2, \ldots, m_N\rangle, \quad \rightarrow \quad \hat{\mathbf{I}}_{jz}|r\rangle = m_j|r\rangle, \tag{2.26}$$

expressing that each product state $|r\rangle$ is a simultaneous eigenstate of the z-component angular momentum operators of all individual spins j. Moreover, these product states are

eigenstates of the weakly coupled spin Hamiltonian (Equation 2.22), with corresponding energy eigenvalues:

$$\omega_r = \sum_j \omega_j^0 m_j + \sum_{j<k} 2\pi J_{jk} m_j m_k. \tag{2.27}$$

In general, the state of a coupled spin system does not coincide with a single energy eigenstate. Instead, arbitrary spin states are described as superpositions of the energy eigenstates, defined in the usual way:

$$|\psi\rangle = \sum_{r=1}^{2^N} c_r |r\rangle, \tag{2.28}$$

where c_r are normalized complex numbers $\sum_{r=1}^{2^N} |c_r| = 1$.

In nuclear magnetic resonance theory, the density operator provides a complete description of the quantum state of an ensemble of spin systems. It is defined as

$$\hat{\rho} = \overline{|\psi\rangle\langle\psi|}, \tag{2.29}$$

where $|\psi\rangle$ denotes the quantum state of an individual spin system, and the overbar indicates an ensemble average over all members of the ensemble [75].

When expressed in the eigenbasis of the spin Hamiltonian, the density operator $\hat{\rho}$ takes a matrix representation whose diagonal elements correspond to the populations of the spin energy levels. These populations are real and non-negative, lie in the interval $[0, 1]$, and are normalized such that their sum equals unity.

The population distribution among the Zeeman energy levels determines the longitudinal net magnetization, which can be expressed in terms of the expectation value of the z-component of the spin angular momentum operator:

$$\mathcal{M}_z \sim \langle \hat{\mathbf{I}}_z \rangle = Tr\{\hat{\rho}\hat{\mathbf{I}}_z\}. \tag{2.30}$$

The off-diagonal elements of the density operator are referred to as *coherences*. These may be classified according to the change in the total z-component of the spin angular momentum,

$$M_r = \sum_j m_j, \tag{2.31}$$

between the two states they connect. A coherence between states $|r_1\rangle$ and $|r_2\rangle$ is said to be of order n if

$$\Delta M = M_{r_1} - M_{r_2} = n. \tag{2.32}$$

In particular, second-order coherences correspond to $\Delta M = \pm 2$.

Only first-order coherences contribute to the transverse magnetization $\mathcal{M}_x$. Moreover, in conventional NMR detection schemes, only single-quantum coherences with $\Delta M = -1$ are observable in the spectrometer. Each such (-1)-quantum coherence is associated with a single spectral transition and therefore gives rise to an individual resonance line in the observed spectrum.

2.4.3 Simple NMR experiment in quantum framework

In this section, the quantum-mechanical description underlying a simple pulsed NMR experiment is briefly and qualitatively outlined. A comprehensive and formal treatment

is beyond the scope of the present thesis and can be found in standard textbooks on NMR spectroscopy (see Ref. [21, 72]).

If the quantum state of the spin ensemble is specified at a given time, its subsequent temporal evolution can be determined by propagating each individual spin state according to the Schrödinger equation:

$$\frac{d}{dt}|\psi\rangle(t) = -\frac{i}{\hbar}\hat{\mathcal{H}}(t)|\psi\rangle(t). \tag{2.33}$$

At the beginning of the experiment, a spin system that has remained undisturbed for a sufficiently long time while interacting with its molecular environment is expected to be in thermal equilibrium with that environment. According to quantum statistical mechanics, a system in thermal equilibrium at temperature T exhibits the following properties:

1. all coherences between the energy eigenstates vanish:

$$\rho_{rs}^{eq} = 0, \quad r \neq s, \tag{2.34}$$

2. the populations of the energy levels follow the Boltzmann distribution:

$$\rho_{rr}^{eq} = \frac{\exp\{-\frac{\hbar\omega_r}{k_B T}\}}{\sum_s \exp\{-\frac{\hbar\omega_s}{k_B T}\}}, \tag{2.35}$$

   where ω_r and ω_s are the transition frequencies associated with states r and s respectively, k_B is the Boltzmann constant, and T is the temperature.

Because the energy separation between Zeeman eigenstates is several orders of magnitude smaller than the thermal energy, the resulting equilibrium population difference is extremely small, allowing the density operator to be approximated as

$$\hat{\rho}^{eq} \simeq \frac{1}{2^N}\left(\hat{1} - \mathbb{B}\hat{\mathbf{I}}_z\right), \tag{2.36}$$

where $\mathbb{B} = \frac{\hbar\gamma B_0}{k_B T}$, and $\hat{1}$ is the identity operator.

The spin precession at thermal equilibrium is expressed in the quantum framework by the effect of a rotation operator acting on the spin states for a time interval τ:

$$\hat{R}_z(\phi) = \exp\{-i\phi\hat{\mathbf{I}}_z\}, \quad \text{where} \quad \phi = \omega_0\tau, \tag{2.37}$$

where ω_0 is the Larmor precession frequency and τ is the duration of the time interval over which the rotation is applied.

In the presence of radio-frequency pulses the spin Hamiltonian becomes time dependent, complicating the equations of motion. This difficulty can be addressed by transforming to a reference frame rotating about the z-axis at a frequency equal to that of the resonant component of the radio-frequency pulse; in this frame, and under suitable approximations, the Hamiltonian becomes effectively time independent and the radio-frequency field appears static.

This transformation can be applied to spin states, the Hamiltonian, and the density operator by means of the unitary operator $\hat{\mathbf{R}}_z(-\Phi)$, where $\Phi(t) = \omega_{\text{ref}}t + \phi_{\text{ref}}$, with ω_{ref} being the reference frequency of the rotating frame and ϕ_{ref} a constant phase offset:

$$|\tilde{\psi}\rangle = \hat{\mathbf{R}}_z(-\Phi)|\psi\rangle; \quad \tilde{\hat{\mathcal{H}}} = \hat{\mathbf{R}}_z(-\Phi)\hat{\mathcal{H}}\hat{\mathbf{R}}_z(\Phi); \quad \tilde{\hat{\rho}} = \hat{\mathbf{R}}_z(-\Phi)\hat{\rho}\hat{\mathbf{R}}_z(\Phi). \tag{2.38}$$

The precession of the spins in the rotating frame will be characterized by the frequency $\Omega^0 = \omega_0 - \omega_{ref}$.

Consider the application of a strong, short radio-frequency pulse, with a flip angle $\beta_p = \omega_{\text{nut}}\tau_p$, where the nutation frequency ω_{nut} is proportional to the amplitude of the radio-frequency field and τ_p denotes the pulse duration. The pulse is assumed to be sufficiently brief that resonance offsets and scalar couplings do not induce any significant evolution during its action. Under these conditions, the spin dynamics during the pulse are dominated exclusively by the radio-frequency interaction:

$$\hat{\mathcal{H}}_p \simeq \omega_{\text{nut}} \left(\hat{I}_x \cos\phi_p + \hat{I}_y \sin\phi_p \right), \tag{2.39}$$

where ϕ_p is the phase of the radio-frequency pulse, which determines the axis in the transverse plane about which the spin rotation is applied. The net effect of the pulse is then propagated by the rotation operator

$$\hat{R}(\phi_p, \beta_p) = \exp\left(-i\hat{\mathcal{H}}_p\tau_p\right). \tag{2.40}$$

Once the radio-frequency pulse is switched off, the spin system undergoes free evolution governed by the internal Hamiltonian. This evolution is described by the free evolution operator, or propagator,

$$\hat{U}(\tau) = \exp\left(-i\hat{\mathcal{H}}^0\tau\right), \tag{2.41}$$

where the static Hamiltonian is given by

$$\hat{\mathcal{H}}^0 = \sum_j \Omega_j^0 \hat{I}_{jz} + \sum_{j<k} 2\pi J_{jk}\, \hat{I}_{jz}\hat{I}_{kz}. \tag{2.42}$$

During free evolution, the coherences acquire a time-dependent phase and thus rotate in the complex plane at a rate determined by the energy separation of the states they connect. In general, the equation of motion for a coherence between states r and s can be written as

$$\rho_{rs}(\tau) \simeq \rho_{rs}(0)\, \exp[-i(\Omega_r - \Omega_s)\tau], \tag{2.43}$$

where $\Omega_r - \Omega_s$ defines the characteristic precession frequency of the coherence. The frequencies of the observed spectral peaks correspond to the precession frequencies of the (-1)-quantum coherences in the rotating frame, while the complex amplitude of each peak is determined by the magnitude of the associated coherence at the onset of the detection period.

Chapter 3

Interpreting the spectra

From manual annotation to automatic analysis

The analysis of one-dimensional ^{1}H Nuclear Magnetic Resonance (NMR) data consists of a sequence of signal processing and interpretation steps that transform raw time-domain measurements into chemically meaningful information. Although the overall workflow is well established in routine NMR practice, individual implementations vary substantially across software packages and applications, and no single standardised procedure exists that is universally adopted across experimental contexts.

For the purpose of reviewing automated and data-driven approaches to NMR analysis, it is therefore useful to delineate a representative sequence of processing and analysis steps and to identify the points at which algorithmic choices directly support the extraction of spectral information. At a high level, the workflow can be divided into three conceptually distinct stages:

1. time-domain signal processing,
2. frequency-domain spectral correction,
3. signal identification and interpretation.

While the present doctoral work is mainly concerned with the analysis stage in which spectral features are detected, characterised, and related to underlying molecular properties, the quality and reliability of the resulting information are strongly influenced by earlier preprocessing and correction steps. These operations determine whether resonances are clearly defined, well separated from background contributions, and consistently represented, thereby affecting the effectiveness of both manual inspection and algorithmic analysis. Accordingly, the following subsection outlines the essential processing operations required to produce spectral representations that preserve the observable characteristics of resonances relevant for downstream analysis.

3.1 Time-domain signal processing

Prior to Fourier transformation, the discrete time-domain signal is often extended by appending additional zero-valued points, a procedure known as zero filling. This operation entails no modification of the measured data and therefore does not increase the true spectral resolution. Nevertheless, by increasing the number of sampled points in the

frequency domain, zero filling yields smoother and more clearly defined spectral lines, which is beneficial for visualization and subsequent data analysis.

The acquisition of a free induction decay inevitably includes noise in addition to the coherent signal. This noise arises most notably from thermal fluctuations in the receiver coil and electronic noise in the radio-frequency hardware. In the time domain, these contributions are well approximated by a zero-mean Gaussian random process. Noise is subsequently transformed into the frequency domain by Fourier transformation. Owing to the linearity of this operation, the statistical properties of the noise are preserved. The noise amplitude in the frequency domain is determined by both the time-domain noise level and the acquisition duration. Reducing the acquisition time lowers the amount of noise accumulated during signal acquisition. However, if the acquisition is truncated too early, significant portions of the free induction decay are lost, leading to signal distortion in the frequency domain. The acquisition time must therefore be chosen to balance noise reduction with accurate representation of the NMR signal.

Furthermore, the initial portion of the free induction decay carries most of the coherent signal, whereas later points are increasingly dominated by noise. To improve the signal-to-noise ratio, it is common practice to apply a decaying weighting function in the time domain, for example

$$w(t) = \exp(-R_{\mathrm{LB}}\, t)\,, \tag{3.1}$$

where R_{LB} controls the decay rate.

Exponential weighting accelerates the apparent decay of the free induction decay and consequently broadens spectral lines in the frequency domain. Because line broadening reduces peak height, overly aggressive weighting can degrade the signal-to-noise ratio despite suppressing late-time noise. The decay parameter must therefore be selected to balance noise attenuation against spectral resolution and signal intensity. It can be shown that the signal-to-noise ratio is maximized when the additional line broadening introduced by the weighting function matches the intrinsic linewidth of the unweighted spectrum; this choice corresponds to a matched filter.

Alternatively, exponential weighting can be combined with a Gaussian function to enhance spectral resolution. Gaussian weighting reduces the long tails characteristic of Lorentzian line shapes and effectively transforms them into approximately Gaussian profiles, whose faster decay improves peak separation and suppresses long-range signal overlap. In practice, this transformation is often incomplete, resulting in spectral lines that are well described by a Voigt profile, reflecting the combined Lorentzian and Gaussian contributions.

3.2 Frequency-domain spectral correction

After Fourier transformation, the spectrum is represented in the frequency domain as a complex-valued signal. Owing to the lack of a fixed phase alignment between the receiver channels and the RF pulse reference frame, an arbitrary and generally unknown phase offset is introduced during signal detection. As a result, neither the real nor the imaginary component of the spectrum corresponds to a pure absorption or dispersion lineshape; instead, each contains a phase-dependent admixture of both contributions:

$$\begin{aligned} S(\Omega) &= S_0[A(\Omega) + i\, D(\Omega)]e^{i\phi} = \\ &= S_0[A(\Omega)\cos\phi - D(\Omega)\sin\phi] + \\ &+ i\, S_0[A(\Omega)\sin\phi + D(\Omega)\cos\phi] \end{aligned} \tag{3.2}$$

Phase distortions are commonly decomposed into two components, referred to as zero-order and first-order phase errors, such that the total phase shift can be expressed as

$$\phi(\Omega) = \phi_0 + \phi_1 \, \Omega. \tag{3.3}$$

Phase correction may be performed manually by adjusting these parameters to obtain spectra in which resonances in the real channel exhibit predominantly absorptive lineshapes. Alternatively, automatic approaches have been developed, such as the ACME algorithm [76], which determines optimal phase parameters by minimising a cost function that combines the entropy of the first derivative of the real spectrum with a penalty term for negative intensities. More recently, data-driven methods based on deep learning have also been introduced, including the `apbk` routine implemented in Bruker TopSpin software.

This routine simultaneously corrects baseline distortions. Baseline correction aims to remove slowly varying background contributions arising from instrumental drift, digitization artefacts, or residual solvent signals, which manifest as low-frequency distortions in the spectrum. An inaccurate baseline can obscure weak resonances and bias intensity-based measurements, rendering baseline correction a critical preprocessing step for automated analysis pipelines.

3.3 Spectral analysis

Once a corrected frequency-domain spectrum is obtained, the analysis transitions from spectral processing to interpretation, in which the information encoded in spectral resonances is mapped onto underlying physical and chemical properties of the system.

The interpretation workflow typically begins with the identification and localization of individual resonance signals, followed by their grouping into multiplets arising from the same nucleus or from sets of magnetically equivalent nuclei. Each multiplet is characterized in terms of its position, intensity, and internal splitting structure, enabling estimation of the number of contributing protons and analysis of scalar coupling patterns. When the molecular formula is known, these features provide the basis for assigning resonance groups to specific atomic sites.

In practice, spectral interpretation must also account for signals that do not originate from the compound of interest. Chemical shifts depend on the solvent environment, and commonly used deuterated solvents are not isotopically pure, giving rise to residual solvent resonances. Additional signals may arise from trace amounts of water or other impurities present in the sample. A robust analysis must therefore distinguish resonances associated with the target molecule from those due to solvent, water, or experimental artefacts.

Beyond individual assignments, the distribution of resonances across the chemical shift axis provides structural constraints, as characteristic functional groups populate specific spectral regions. This information can be combined to propose candidate molecular structures in a process of structure elucidation. These candidates may subsequently be evaluated through structure verification, by predicting the corresponding NMR spectrum and comparing it with the experimental data. Together, these steps define the overarching workflow of spectral interpretation, which is elaborated in detail in the following subsections.

3.3.1 Peak picking

The analytical step aimed at extracting a discrete representation of a spectrum in terms of individual resonance positions is commonly referred to as peak picking. Traditional peak picking methods typically rely on the identification of local extrema, analysis of spectral derivatives, application of noise or intensity thresholds, and exploitation of geometric or symmetry properties of resonance lines [22]. To limit spurious detections, several algorithms operate on transformed representations of the spectrum prior to peak picking. WavPeak [77] applies a wavelet transform to enhance signal features across multiple scales. Non-negative matrix factorization [24] approximates the spectrum as a sum of separable one-dimensional components, while the PICKY algorithm [25] uses singular value decomposition to extract orthogonal spectral contributions. In its simplest form, peak picking produces a list of peak frequencies, but in most applications it is desirable to also estimate additional peak attributes such as amplitude, linewidth, and line shape. This typically requires modeling the spectrum as a superposition of parametric components, for example Lorentzian, Gaussian, or Voigt profiles, and therefore entails an implicit deconvolution of the measured signal. In a Bayesian framework [30] the spectrum can be modeled as a mixture of Gaussian peaks, and infer peak parameters via stochastic sampling.

Recovering an exact deconvolution of an NMR spectrum is, however, an ill-posed problem. Multiple distinct combinations of peak parameters can reproduce the observed data within a comparable residual error, implying that the solution is not unique. In particular, minimizing the discrepancy between the reconstructed and experimental spectra alone can lead to pathological solutions in which an excessive number of peaks is introduced to account for noise, baseline imperfections, or minor spectral distortions. Such solutions are mathematically admissible but lack physical or practical relevance.

Selecting a meaningful deconvolution therefore requires additional criteria beyond residual minimization. This is commonly achieved by incorporating prior assumptions or constraints on the solution, for example limiting the number of components, restricting parameter ranges, or penalizing excessive model complexity. These forms of regularization guide the deconvolution towards solutions that balance fidelity to the data with robustness and interpretability, yielding peak lists that are suitable for subsequent spectral analysis.

The CRAFT (Complete Reduction to Amplitude-Frequency Table) procedure is a representative example of a traditional fitting method that operates directly on FID data. In contrast, a broad class of alternative techniques exploits the autoregressive behavior of the FID signal, including Hankel matrix–based approaches, filter diagonalization, and fast Padé transform methods. In these methods, each new point in the FID is expressed as a linear combination of a set of preceding points. A different strategy works entirely in the frequency domain, such as the Global Spectral Deconvolution (GSD) [29] method proposed by Cobas and co-workers, which performs deconvolution by using the spectrum together with its first- and second-order derivatives for peak detection and line-shape fitting.

A recent example of probabilistic peak picking is UniDecNMR [31], which is based on a Bayesian deconvolution framework originally developed for mass spectrometry data, a problem sharing several characteristics with NMR spectral analysis. In this approach, the measured intensity spectrum is modeled as the convolution of a peak shape function with a sparse array of sources, represented as delta functions over the spectral frequency grid. The deconvolution removes the explicit line shape contribution, yielding a source

distribution that directly encodes peak positions and intensities. In practice, the method relies on an accurate, user-defined specification of the line shape parameters, typically estimated from an isolated resonance prior to analysis, and of the amplitude threshold.

Deep learning approaches, by contrast, aim to reduce or eliminate manual parameter tuning by learning peak representations directly from data and producing predictions in an end-to-end fashion. A representative example is NMRNet [45], which casts peak detection as a binary classification problem on localized two-dimensional spectral patches analyzed by a convolutional neural network. In contrast, DEEP picker [46] adopts a feature-extraction strategy based on one-dimensional convolutions applied along each spectral axis, enabling the prediction not only of peak positions but also of amplitudes and line shapes, and supporting spectra of arbitrary dimensionality. More recently, deep residual networks have been integrated into fully automated pipelines, most notably within the ARTINA framework [47], where a two-stage neural architecture combines coarse peak localization with regression-based refinement. Trained on large collections of experimentally acquired multidimensional protein NMR spectra, ARTINA demonstrates robust, end-to-end peak detection without user intervention and is supported by publicly available benchmark datasets.

Recently MR-Ai [48], a physics-informed framework, leverage the specific point spread function (PSF) of the sampling schedule to map experimental data directly to peak probability distributions. This approach is implemented by training a deep neural network, inspired by the WaveNet architecture [78], to recognize the global interference patterns and aliasing artifacts inherent in non-uniform sampling (NUS) schedules. The model transforms traditional intensity-based signals into a statistical "Peak Probability Presentation" (P^3), effectively converting the peak detection task into a probability mapping that can resolve highly overlapped resonances.

mldcon in Bruker Topspin While most deep learning approaches have focused on multidimensional protein NMR, neural methods have also been proposed for one-dimensional spectra of small molecules. An hybrid architecture [49] for the deconvolution of proton NMR spectra was developed within an Innosuisse-funded collaboration with Bruker, a leading manufacturer of NMR spectrometers and associated analysis software. This work was carried out within the same collaborative framework in which the multiplet classification algorithm described in the following chapters was developed. The study reformulated peak detection and parameter estimation as a joint classification–regression task aimed at disentangling overlapping resonances in complex spectra. A translation-invariant neural network was trained on large libraries of synthetically generated spectra with known ground-truth peak parameters, thereby avoiding the need for manual annotation of experimental data. The model assigns local spectral regions to baseline, narrow, or broad resonances while simultaneously predicting amplitudes and linewidths, using a labeling strategy that accounts for signal-to-noise variations. When evaluated on experimental spectra of small molecules, the approach demonstrated robust performance in crowded regions, across large intensity disparities, and in the presence of partially overlapping peaks, while maintaining a low false-positive rate. Owing to this level of reliability and practical applicability, the algorithm has subsequently been incorporated into the commercial Bruker analysis software TopSpin, under the command mldcon.

3.3.2 Splitting pattern detection

Beyond the identification of resonance positions and intensities, a central step in the interpretation of one–dimensional ^{1}H NMR spectra is the analysis of splitting patterns arising from scalar spin–spin couplings. These patterns encode quantitative information on the number, type, and relative arrangement of magnetically coupled nuclei, and their accurate interpretation enables the estimation of scalar coupling constants. In the interpretation of one–dimensional ^{1}H NMR spectra, information derived from the analysis of coupling–induced splitting patterns is often more robust than that obtained from chemical shift values alone. Chemical shifts are strongly influenced by local electronic environments and cannot, in general, be predicted with high precision from first principles, nor reliably transferred across nominally similar structural motifs. By contrast, scalar coupling patterns obey well–defined spin–interaction rules and exhibit comparatively limited variability. As a consequence, when structural inferences drawn from chemical shifts and from splitting patterns are inconsistent, coupling–based analysis is commonly regarded as the more reliable source of information [6]. From a practical standpoint, splitting pattern analysis consists in identifying sets of equally spaced resonances within a multiplet and across related multiplets, from which scalar coupling constants can be quantitatively extracted. Importantly, scalar couplings are reciprocal: if a given multiplet exhibits a splitting corresponding to a coupling with another resonance, the same coupling constant must manifest as an equivalent spacing in the partner multiplet. The consistent occurrence of identical peak separations across coupled peak sets therefore provides a self–consistency criterion that can be exploited to validate and refine the extraction of coupling constants.

In experimental spectra, however, the direct extraction of coupling parameters is often hindered by peak overlap, limited spectral resolution, relaxation–induced line broadening, and noise, which distort or partially mask the ideal multiplet structures predicted by first–order theory. As a result, splitting pattern analysis has motivated the development of specialized algorithms that move beyond isolated peak detection and instead aim to identify, model, or decompose multiplets as structured entities from which coupling constants can be inferred in a robust and reproducible manner.

Similar to what occurred in the development of automated peak picking, early attempts at retrieving multiplet splitting patterns and scalar coupling constants from one-dimensional NMR spectra relied on explicit signal models and handcrafted heuristics. These approaches reflected both the limited computational resources available at the time and a strong dependence on analytical descriptions of resonance line shapes and coupling topologies. Initial methods exploited local regularities of resonance patterns, such as symmetry relations within multiplets, to infer coupling structures and peak groupings, as exemplified by the work of Boentges et al. (1989) [23]. In parallel, information–theoretic formulations based on maximum entropy principles were introduced to regularize ill-posed inverse problems arising in spectral reconstruction and deconvolution. Studies by Delsuc and Levy (1988) [26], and later by Seddon et al. (1996) [27] and Stoven et al. (1997) [28], demonstrated that entropy-based constraints can stabilize the estimation of peak positions and amplitudes in the presence of noise and overlap, albeit at the cost of nontrivial parameter tuning and sensitivity to prior assumptions.

A complementary line of research focused on explicit multiplet deconvolution, treating each resonance as a superposition of parametric line shapes whose relative positions and intensities encode scalar coupling information. An iterative deconvolution approach for

extracting J-coupling values was introduced by Jeannerat and Bodenhausen (1999) [32]. This procedure operates by applying a sequence of equidistant delta functions across the spectral data. The precise coupling constant is determined by systematically adjusting the intervals between these functions until the resulting spectral profile is optimally simplified. However, a significant limitation of this methodology is the concurrent amplification of noise and spurious artifacts along the axis of deconvolution. This signal degradation becomes distinctly more problematic when attempting to resolve very narrow splittings. Consequently, to minimize these detrimental effects, the algorithm must be strictly confined to a carefully isolated spectral window that exclusively encompasses the multiplet of interest. Pattern-recognition strategies were explored by Golotvin and Chertkov (1997) [34]. Their algorithm relied extensively on the accurate definition of an initial theoretical model for the resonance at hand. This approach remained tightly coupled to handcrafted descriptors and exhibited limited robustness to deviations from idealized line shapes.

A particularly influential contribution was the recursive splitting-tree algorithm introduced by Hoye and co-workers [35, 36], which formalized multiplet interpretation as a hierarchical decomposition governed by scalar couplings. To function effectively, this procedure fundamentally relied on the multiplet exhibiting the symmetrical peak distributions and characteristic intensity ratios inherent to strictly first-order spectra. Moreover, the algorithm required explicit prior inputs from the user: the total number and specific frequencies of the constituent peaks, along with the quantity of active scalar couplings. Subsequent extensions of he Hoye algorithm incorporated additional symmetry relations and amplitude constraints [37, 38], improving performance robustness on real experimental spectra but also leading to suboptimal estimate of the J-coupling in complex cases. Nevertheless, these algorithms were intrinsically local: they operated on isolated resonances and assumed prior knowledge ofmultiplicity.

To overcome the resonance-by-resonance limitation, Griffiths [39] proposed a global aggregation strategy in which individual peaks are first detected and subsequently clustered into multiplets across the entire spectrum. Although this represented a conceptual step toward full automation of multiplet identification, the procedure still relied on accurate, externally provided peak positions, rendering the pipeline sensitive to upstream errors in peak picking and phase correction.

Despite these limitations, classical model-based methodologies remain in active use, particularly within expert-driven workflows and commercial software [33]. However, their reliance on explicit assumptions about spectral structure, noise statistics, and resonance multiplicity has motivated a progressive shift toward data-driven approaches. In this context, recent deep learning applications have largely focused on the prediction of chemical shifts and scalar coupling constants from molecular structure, using supervised machine–learning models trained on quantum–chemical reference datasets. Graph–based neural networks and kernel regression approaches have demonstrated high accuracy in this forward–prediction setting, but they presuppose prior knowledge of the molecular structure and do not analyze experimental NMR spectra [50, 51]. On the other hand, the inverse problem of extracting coupling parameters directly from one–dimensional ^{1}H NMR spectra remains comparatively less explored in the literature, despite its central importance in spectral analysis.

3.3.3 Going from the spectrum to the molecule

A first step in extracting structural information from a ^{1}H NMR spectrum is the assignment of each resonance to a group of magnetically equivalent protons. This assignment is inferred from the signal's position along the chemical shift axis together with its splitting pattern and scalar coupling constants, which collectively constrain its local environment and spin connectivity. Several commercial software packages have implemented automated assignment algorithms that formalize this reasoning computationally. These systems combine rule based spectral interpretation with database matching and statistical scoring schemes to assist or partially automate the mapping between observed resonances and candidate proton environments. In more recent implementations, this workflow is augmented by machine learning models, in particular deep neural networks trained to predict chemical shifts from molecular structure, which provide prior estimates that can be integrated into the assignment procedure. Cordova et al. [70] combined a Bayesian probabilistic framework with deep learning derived chemical shift predictions, using the latter to define informative priors over expected resonance positions for distinct covalent environments. Within this formulation, experimental shifts are integrated with the learned priors through Bayesian inference to yield statistically weighted assignments, thereby coupling data driven prediction with principled uncertainty quantification.

Recently, NMRformer [79] was introduced as a transformer based framework for peak assignment and metabolite identification in one dimensional ^{1}H NMR spectra of complex mixtures. Validated on multiple cellular and biofluid datasets, the method achieved peak assignment accuracies above 88% and metabolite identification accuracies above 80%, demonstrating robust performance across diverse sample types.

For structurally simple compounds, the molecular structure can in some cases be inferred directly from the ^{1}H NMR spectrum in combination with the molecular formula. The relative integrals of resolved resonance groups provide access to the number of contributing protons, while their chemical shifts constrain the associated functional environments. Characteristic splittings and recurring coupling patterns further allow the identification of adjacent proton groups, enabling the progressive assembly of local structural motifs into larger substructures. By iterating this reasoning, one may arrive at one or a small set of plausible candidate structures, a process commonly referred to as structure elucidation.

As molecular complexity increases, however, this approach rapidly loses reliability. Spectral overlap becomes prevalent, integrals deviate from idealized integer ratios, and individual resonance patterns are no longer unambiguously assignable. In such cases, the combined analysis of ^{1}H and ^{13}C spectra together with correlation experiments such as DEPT, COSY, HSQC, HMQC, and HMBC enables the reconstruction of the underlying molecular connectivity. Beyond this multidimensional strategy, recent work [80] has demonstrated that molecular structure can also be inferred directly from routine one dimensional ^{1}H and or ^{13}C spectra using end to end learning architectures. Spectral features are extracted via a convolutional neural network and processed by a transformer model that assembles molecular fragments into coherent graph structures, yielding exact structure predictions in 69.6% of cases within the first 15 ranked candidates and reducing the effective search space by up to eleven orders of magnitude. In a complementary formulation, DeepSPInN [81] casts structure elucidation from infrared and ^{13}C NMR spectra as a Markov decision process solved through deep reinforcement learning with Monte Carlo tree search, enabling systematic exploration of chemical space without reliance on spec-

tral libraries or manually encoded heuristics. However, even when a consistent structure is obtained, independent verification remains advisable, for instance through comparison with spectra predicted from candidate structures.

Chapter 4

Learning signal phenotypes

4.1 Multiplet splitting pattern as a segmentation problem

Following spectral acquisition, the subsequent analysis aims at extracting structural information from the observed resonances. As introduced in the previous chapter, peak identification can be formulated as a supervised pointwise binary classification task [46, 49]. In this setting, each sampled point of the spectrum is assigned a label: 1 for a peak, indicating the presence of a genuine NMR signal, and 0 for non-peak regions, corresponding to baseline fluctuations or experimental artefacts. However, spectral analysis often requires a deeper level of interpretation beyond simple peak detection. In particular, individual resonances rarely appear as isolated lines in ^{1}H NMR: through-bond spin–spin interactions give rise to characteristic splitting patterns, or multiplets, whose internal structure encodes information on the number, identity, and coupling strengths of neighboring nuclei. Correctly resolving and grouping these correlated spectral components is therefore essential for an accurate interpretation of NMR spectra and cannot be addressed by peak detection alone.

In this context, the classification task can be extended from the detection of individual spectral points to the identification of entire multiplets, with each splitting pattern (e.g., singlet, doublet, triplet) assigned a distinct integer label. The absence of a signal can remain labeled as 0. We therefore developed a deep learning algorithm that outputs, for each sampled spectral point, a prediction of the corresponding splitting class. In this formulation, the spectral axis is effectively partitioned into contiguous regions corresponding either to baseline or to resonances associated with distinct multiplet types. From a machine-learning perspective, this corresponds to a one-dimensional semantic segmentation task [82], analogous to the assignment of class labels to pixels for the purpose of recognizing and localizing objects in an image.

Obtaining a fully segmented spectral range provides an effective and scalable automatic annotation strategy, which serves multiple purposes:

Classification: The segmentation explicitly separates baseline regions from resonant signals, while simultaneously assigning each detected resonance to a specific splitting class.

Localization: By assigning labels at the level of individual spectral points, the method yields precise estimates of both the chemical shift position and the spectral extent

of each resonance or multiplet, facilitating accurate signal boundary determination.

Quantitative analysis: The resulting segmented representation provides a structured input for downstream quantitative tasks, including signal integration, estimation of coupling constants from multiplet structure, and subsequent stages of automated spectral interpretation.

4.1.1 Network architecture

To process one-dimensional spectra, we employed a Convolutional Neural Network (CNN). The proposed architecture [52, 49, 83] is organized into two key computational stages (see Figure 4.1 and Table 4.1): the initial feature extractor block processes the input data into meaningful discriminative representations, while the subsequent feature correlation block analyzes relationships between the extracted features, identifying and quantifying interdependencies. The feature extractor block processes the input spectrum through an inception module [84] containing four parallel 1D convolutional layers with progressively larger kernel sizes (ranging from 4 to 256 points). All convolutional layers were followed by Exponential Linear Unit (ELU) activation functions to introduce nonlinearity into the feature transformations. This multi-scale architecture enables the network to capture features across different spectral ranges, which is particularly important for NMR data where resonances exhibit varying widths. Specifically, signal patterns (e.g., singlets vs multiplets) span different numbers of spectral points depending on their splitting patterns and physical parameters (e.g., coupling constants and linewidths). The extracted features are concatenated and fed into the correlation block. First, a time-distributed fully connected layer processes the features, followed by a bidirectional Long Short-Term Memory (LSTM) layer. The bidirectional architecture enables information flow in both spectral directions through the network's recurrent modules. This design proved crucial for our application, as it ensures that predictions for each spectral point incorporate contextual features from both upfield and downfield regions while maintaining classification invariance to multiplet positions along the frequency axis.

The subsequent architecture comprises four time-distributed dense layers with progressively decreasing dimensionality (from 512 nodes to the number of output classes), each employing Rectified Linear Unit (ReLU) activation. The network culminates in a softmax output layer that generates, for each spectral point, a probability distribution across all possible multiplet classes.

During the training phase, the network processed input spectral samples ($\mathbf{x}$) along with their corresponding point-wise labels ($\mathbf{y}$). All 1,225,606 trainable parameters ($\boldsymbol{\theta}$) were optimized at each epoch through backpropagation to minimize the Focal loss [85] objective function. Focal loss is a modification of the cross-entropy loss designed to mitigate class imbalance by down-weighting well-classified examples and concentrating learning on hard, misclassified ones. In the multiclass setting, the modulating factor is applied to the cross-entropy term corresponding to the true class only. For a single spectral point with an integer label $y \in \{0, \ldots, K-1\}$ and a corresponding predicted probability p_y, the multiclass focal loss is defined as

$$\mathcal{L}(y, \mathbf{p}) = -(1 - p_y)^{\gamma_{fl}} \log(p_y), \tag{4.1}$$

where $\gamma_{fl} \geq 0$ is the focusing parameter. The loss reduces to the standard cross-entropy for $\gamma_{fl} = 0$.

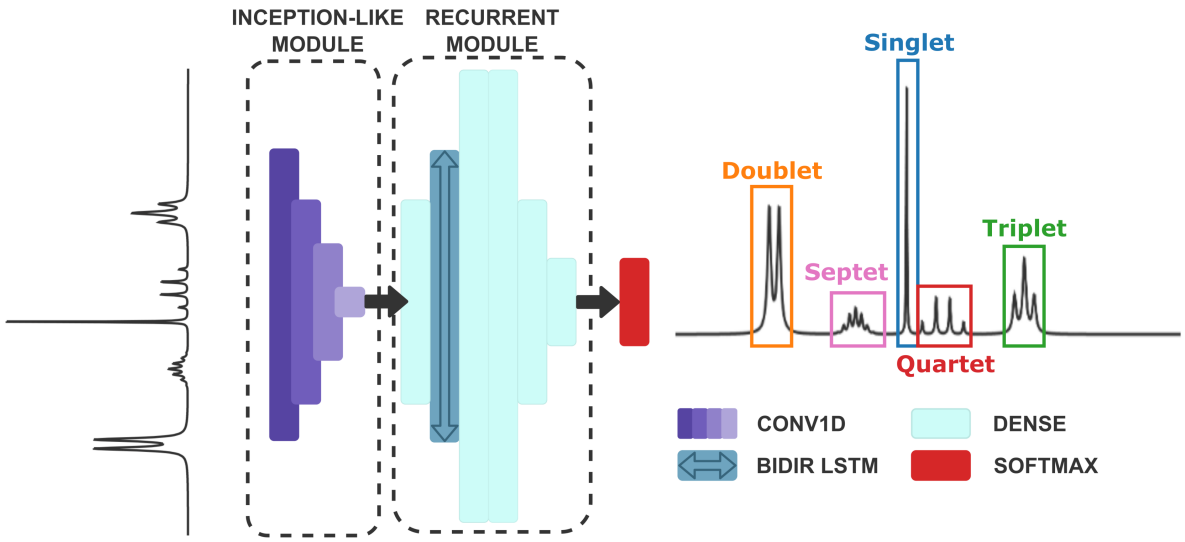


Figure 4.1: Diagram of the basic architecture.

For $\gamma_{fl} > 0$, the loss progressively suppresses the contribution of easy examples with $\hat{p}_y \approx 1$, while preserving a large gradient for hard or misclassified samples with $\hat{p}_y \approx 0$. A commonly used value is $\gamma_{fl} = 2$, which provides a practical balance between stability and selective emphasis. This property is particularly advantageous in the present segmentation task, where the dataset is inherently imbalanced: baseline points constitute the majority of spectral samples, and their discrimination from signal regions is substantially easier than the finer-grained classification among different multiplet splitting classes. By reducing the dominance of easily classified baseline regions, focal loss encourages the network to allocate greater capacity to resolving spectrally informative but underrepresented classes.

At the end of the training phase, the network had learned a complex mapping function $\Phi_\theta(\mathbf{x})$ that transforms input spectra into multiplet classifications. For prediction, the model outputs a probability distribution over all possible classes via the softmax function, with the final label prediction $\hat{\mathbf{y}}$ determined by selecting the class with maximum probability: $\hat{\mathbf{y}} = \text{argmax}_l \left[\text{Softmax}(f_\theta(\mathbf{x}))_l\right]$, where $l \in \{0, \ldots, K-1\}$ indexes the output classes.

Table 4.1: *Network architecture.* The table reports the output dimensions for each layer, where S_B is the batch size, S_I is the input size, that is the number of points in the spectral segments, and the third component is the channel size.

Layer	Output
Inception-like module	$(S_B, S_I, 64)$
Time-distributed fully connected	$(S_B, S_I, 64)$
Bi-directional LSTM	$(S_B, S_I, 256)$
Time-distributed fully connected	$(S_B, S_I, 512)$
Time-distributed fully connected	$(S_B, S_I, 512)$
Time-distributed fully connected	$(S_B, S_I, 64)$
Time-distributed fully connected	$(S_B, S_I, 8)$
Softmax output	$(S_B, S_I, 8)$

4.2 Training set

Deep learning algorithms require large amounts of training data to avoid overfitting and achieve strong generalization performance. In recent years, the focus of both researchers and practitioners has gradually shifted from architectural innovations to improving the quality and quantity of training data [86]. In addition to quantity, the data must also be information-rich, meaning it should exhibit a high degree of diversity in its feature space. This is because deep models are inherently better at interpolating within the distribution of the training data than extrapolating beyond it. As a result, accurate and meaningful predictions can only be achieved when the model is trained on a sufficiently large and diverse dataset.

However, collecting a real-world ^{1}H NMR spectral dataset that meets these requirements is a demanding task. Assembling a database of sufficient size to train a deep neural network, while simultaneously ensuring adequate coverage of experimental variability, would require a substantial and sustained effort. In particular, such a dataset should effectively sample the chemical space of small molecules, encompassing the relevant classes of spin systems, and should span a wide range of experimental conditions, including temperature, pH, solvent composition, and signal-to-noise ratios. Achieving this level of diversity would necessitate extensive and repeated use of high-field NMR spectrometers, making the acquisition process resource-intensive, both economically and in terms of human effort. Moreover, manual annotation of such a dataset would be extremely demanding and heavily reliant on expert knowledge. Human labelling also introduces variability: different experts may define label ranges inconsistently, and in complex spectral patterns such as fast chemical exchange, severe peak overlap, or the presence of experimental artefacts, the interpretation of spectral features may legitimately differ across experts, potentially undermining the learning process.

In contrast, generating synthetic spectra in silico offers a scalable and customizable alternative. It enables the creation of virtually unlimited training examples and allows for precise tailoring of the dataset to the specific needs of the application. Additionally, the synthetic generation process is coupled with fast, automated, and consistent labelling, replacing months of manual expert work with just a few hours of computation, and ensuring label uniformity across the dataset.

4.2.1 Synthetic data generation

We developed an algorithm to generate synthetic ^{1}H NMR spectral segments for training our deep learning models, based on random sampling of relevant spectral features. This approach does not rely on a quantum-mechanical simulation of the spin system Hamiltonian of the molecular sample. While such physically grounded simulations can offer high fidelity in modeling magnetic interactions and signal behavior, they come with significant computational costs. In particular, simulating the full spin dynamics of complex or large spin systems often becomes prohibitively time-consuming and resource-intensive. Instead, we opted for a pragmatic and efficient strategy, aiming to faithfully reproduce the phenotypical characteristics of resonances as they appear in real ^{1}H NMR spectra. By focusing on the observable patterns, such as peak positions, splitting structures, and lineshapes, rather than the underlying quantum mechanics, our method enables the rapid generation of diverse and realistic spectral segments suitable for training deep learning models at scale. Each segment was randomly assigned a number of multiplets, with their

positions, splitting patterns, and coupling constants drawn from randomized distributions (see Table 4.2).

Table 4.2: Parameter definitions used for synthetic spectrum generation. $Exp(\lambda)$ refers to the Exponential distribution function and $B(\alpha, \beta)$ refers to the Beta distribution function.

Parameter	**Definition / Range**
Spectral width (ppm)	1 ppm
Number of points	2048
Spectrometer frequency	400 MHz
sino	sampled in the range $10^{1/2}$ to $10^{5/2}$
Number of features	$4 + int(\mathrm{Exp}(\lambda = 1.0))$
Atom count	integer in [1, 10]
Chemical shift	uniform over the full spectral range
Lineshape parameter l_s	random in [0, 1]
Gyromagnetic factor γ	$0.5 + 16.5 \cdot \mathrm{B}(\alpha = 2, \beta = 20)$
Scalar coupling J_1	uniform in [0, 20] Hz
Scalar coupling J_2	uniform in $[0, J_1]$
Scalar coupling J_3	uniform in $[0, J_2]$
Scalar coupling J_4	uniform in $[0, J_3]$
Mixing weight w	uniform in [0.45, 0.55]
Zero-order phase ϕ_0	uniform in $[-6, 6]$
First-order phase ϕ_1	uniform in $\left[-\frac{2}{\text{ppm\_range}}, \frac{2}{\text{ppm\_range}}\right]$
Left baseline offset b_l	uniform in $[-2\,\text{noise}, 2\,\text{noise}]$
Right baseline offset b_r	uniform in $[-2\,\text{noise}, 2\,\text{noise}]$

According to NMR theory, peak shapes ideally follow a Lorentzian profile [21]. The intrinsic NMR signal generated by nuclear spin precession has a Lorentzian line shape, reflecting the exponential decay of transverse magnetization governed by the T_2 relaxation time and homogeneous broadening mechanisms. In real experimental conditions, however, deviations from ideality lead to peaks adopting a shape closer to a Voigt profile, defined as the convolution of a Lorentzian and a Gaussian component. A Gaussian contribution arises from inhomogeneous broadening effects, such as static magnetic field inhomogeneities and local magnetic susceptibility variations within the sample. In addition, the application of Gaussian apodization functions during signal processing further enhances the Gaussian character of the observed line shape. This effect can be efficiently modeled using the Pseudo-Voigt approximation [87]:

$$V(\omega - \omega_0; \gamma_w) \equiv (1 - l_s)L(\omega - \omega_0; \gamma_w) + l_s G(\omega - \omega_0; \sigma_g), \tag{4.2}$$

where ω_0 denotes the peak position, L the Lorentzian profile characterized by its Half Width at Half Maximum (HWHM) γ_w, and G represents the Gaussian profile with variance σ_g. To ensure comparability, both lineshapes were generated with matching HWHM, enforcing $\gamma_w \equiv \sigma_g\sqrt{2\log(2)}$. The linewidth parameter γ_w was drawn from a distribution with support between 0.5 and 8 Hz., while the lineshape parameter l_s was varied between 0 (pure Lorentzian) and 1 (pure Gaussian), allowing us to span a realistic range of peak shapes observed in experimental spectra. Linewidths and intensities were also randomly varied within experimentally reasonable ranges.

To simulate second-order effects, we included the rooftop effect, deviating from the ideal Pascal triangle amplitude ratios. The rooftop distortion arises from the coupling strength J between interacting nuclear spins in the strong coupling regime. Various configurations for the resonance of a spin influenced by two distinct coupling constants are illustrated in Figure 4.2. In this example, the distortion may be associated with

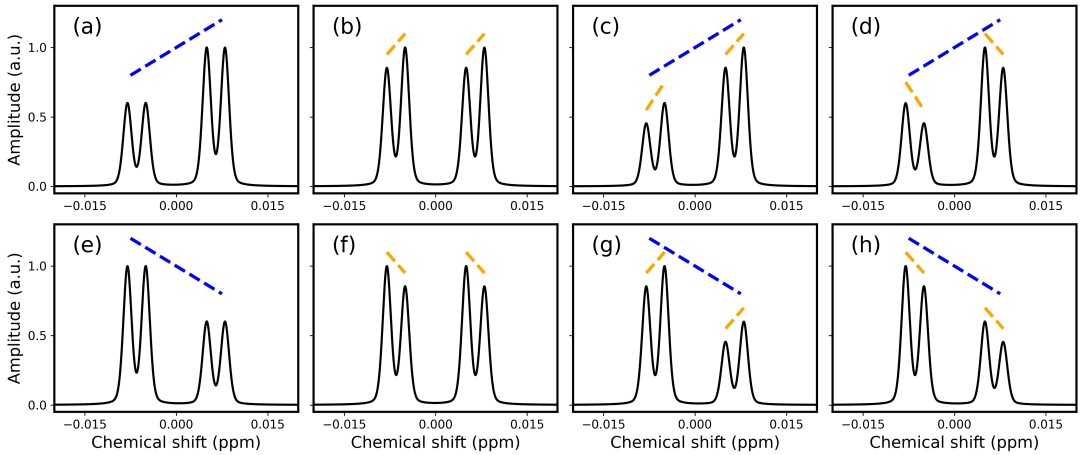


Figure 4.2: All possible rooftop configurations for a doublet of doublets with $J_1 > J_2$. Panels (a, e) and (b, f) illustrate the case in which the rooftop effect is applied to only one coupling (respectively J_1 and J_2). The remaining panels show the full set of configurations obtained when the rooftop effect is applied to both couplings.

either one of the coupling constants individually or with both simultaneously but independently. According to the quantum-mechanical framework of NMR, such distortions typically manifest symmetrically in the signals of both coupled nuclei, producing what is known as a rooftoping couple. In contrast, within our synthetic training set, we applied this effect in a simplified manner by altering the peak intensities of each generated multiplet independently. Operationally, this simplified rooftop distortion was implemented by redistributing the relative intensities of the individual lines within an m-component multiplet according to a single mixing parameter w, constrained to the narrow interval $w \in [0.45, 0.55]$. Specifically, the ideal Pascal-triangle amplitude ratios were replaced by binomially weighted coefficients of the form

$$A_k = \binom{m-1}{k} w^k (1-w)^{m-1-k}, \qquad k = 0, \ldots, m-1, \tag{4.3}$$

which are normalized by construction. Restricting w to values close to $1/2$ ensures that the resulting intensity patterns represent mild, near-symmetric deviations from the ideal first-order case, consistent with weak-to-moderate second-order rooftop distortions. This parametrization preserves the total signal intensity while introducing controlled asymmetry in the multiplet line strengths, providing a pragmatic approximation of rooftop effects without explicitly modeling the strong-coupling spin Hamiltonian.

Our method was tailored for ^{1}H NMR spectra that had already undergone phase and baseline correction. However, in order to improve robustness and promote generalization, we deliberately introduced mild phase and baseline distortions to mimic residual imperfections commonly encountered in experimental data. A frequency-dependent phase distortion was applied according to

$$\Delta\phi(\omega) = \phi_0 + \phi_1\omega, \tag{4.4}$$

Table 4.3: Comparison between prototype network labels and the corresponding MuSe Net multiplet encodings.

Prototype network			MuSe Net		
Numerical label	Symbol	Description	Numerical label	Symbol	Description
0	–	No signal	0	–	No signal
1	s	Singlet	1	s	Singlet
2	d	Doublet	2	d	Doublet
3	t	Triplet	3	t	Triplet
4	q	Quartet	4	q	Quartet
5	quint	Quintet	5	quint	Quintet
6	s	Sextet	6	dd	Doublet of Doublets
7	p	Septet	7	dt	Doublet of Triplets
			8	td	Triplet of Doublets
			9	tt	Triplet of Triplets
			10	ddd	Doublet of Doublets of Doublets
			11	dddd	Doublet of Doublets of Doublets of Doublets

with the parameters constrained to $|\phi_0| \leq 6$ rad and $|\phi_1| \leq 2$ rad. Baseline distortions were introduced by defining two parameters, b_l and b_r, whose absolute values were restricted to be smaller than twice the noise level. The resulting baseline contribution was modeled as

$$\Delta b(x) = b_r\, x + b_l\, (1 - x), \tag{4.5}$$

where x is a linear coordinate increasing monotonically from 0 at the left boundary of the spectrum to 1 at the right boundary. This construction yields a weakly sloped baseline that preserves the overall spectral structure while accounting for realistic residual distortions. Additionally, we added Gaussian noise to cover a range of signal-to-noise ratios ranging between ~ 3 to ~ 300.

Using this synthetic generation strategy, we created two training sets of $100'000$ spectral segments each, corresponding to two stages in the development of our deep learning framework. In the first stage [52], we produced segments containing 1024 points with a spectral resolution of 5 points per Hertz. We focused on multiplets characterized by a single coupling constant. The splitting patterns in each spectrum were randomly selected from singlets, doublets, triplets, quartets, quintets, sextets, and septets. In the second stage [83], we included eleven distinct splitting patterns featuring up to four coupling constants, selected to reflect those most frequently encountered in experimental spectra[1] (see Table 4.3): singlets, doublets, triplets, quartets, quintets, doublets of doublets, doublets of triplets, triplets of doublets, triplets of triplets, doublets of doublets of doublets, and doublets of doublets of doublets of doublets. After generating the multiplets, we labeled the spectra by assigning each point the class value corresponding to its multiplet. The set of multiplet labels is defined as a subset of the natural numbers, $\mathcal{C} \subset \mathbb{N}$. Baseline points, those not belonging to any signal, were assigned the label 0. Table 4.3 provides the mapping between numerical labels and multiplet classes for both training sets.

[1]Based on internal statistics at partner company Bruker Switzerland.

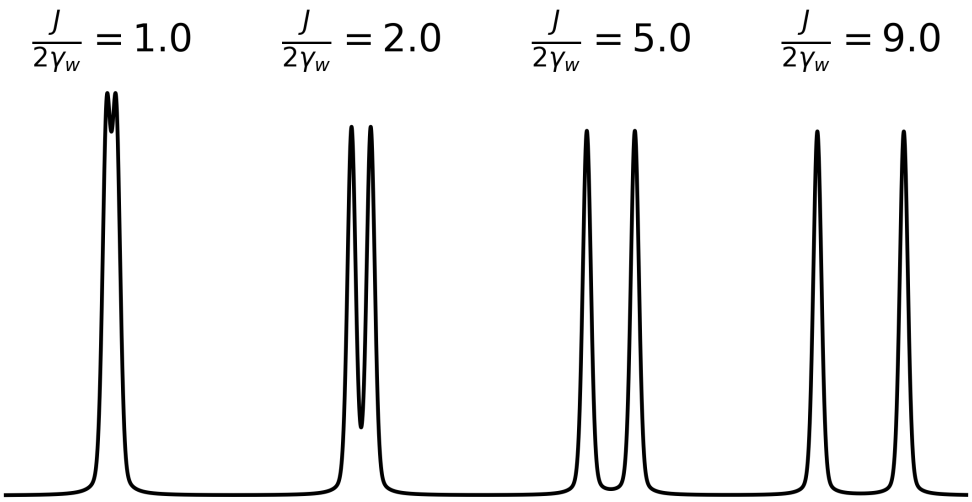


Figure 4.3: Display of different peak resolution regimes for a doublet occurring in the synthetic dataset used in the prototype version of the algorithm. The spectra illustrate increasing values of the dimensionless ratio $\frac{J}{2\gamma_w}$, highlighting the transition from unresolved to well-resolved multiplet structures as the scalar coupling becomes large compared to the homogeneous linewidth.

4.2.2 Regularization mechanisms

Randomly sampling all spectral parameters ensures that the network encounters a wide range of cases during training, thus enhancing its ability to generalize. Nonetheless, to prevent illogical or inconsistent examples, we implemented several regularization mechanisms. Firstly, when the generation process led to overlapping multiplets within a segment, the segment was discarded. The same rule applied if more than 20% of a multiplet's signal area extended beyond the defined spectral range. Additionally, we regularized the features characterizing different resonance patterns. During synthetic generation, certain combinations of coupling constants, peak widths, and intensities caused neighboring peaks to merge, forming broader and distorted signals. As a result, the number of visible peaks might differ from the expected count for a given multiplet class. In one-coupling resonances, if the ratio between the coupling constant and the peak width $J/2\gamma_w$ is small the peaks tend to merge together into a single peak. In resonances with multiple couplings, specific ratios between coupling constants could make one multiplet resemble another. These cases risk confusing the classification algorithm. To address this, we defined a clear multiplet phenotype for each class and included only multiplets with all peaks clearly visible. We adopted two different and specific regularization strategies for the two versions of the training set. In the first version, to ensure that all peaks can be distinguished while representing different levels of peak resolution, we varied the ratio of the distance between consecutive peaks in a multiplet, which for the resonance patterns considered corresponds to the coupling constant J, and the Full Width at Half Maximum (FWHM) of the peaks $2\gamma_w$, between 0.5 and 18 (see Figure 4.3).

On the other hand, for signals with multiple couplings, we also developed a relabeling algorithm to correct multiplet labels when the resulting signal phenotype mimicked a different class (see Figure 4.4).

This relabeling involved peak picking on synthetic spectra after shrinking the linewidth according to the signal-to-noise ratio [49], enhancing the recognition of closely spaced

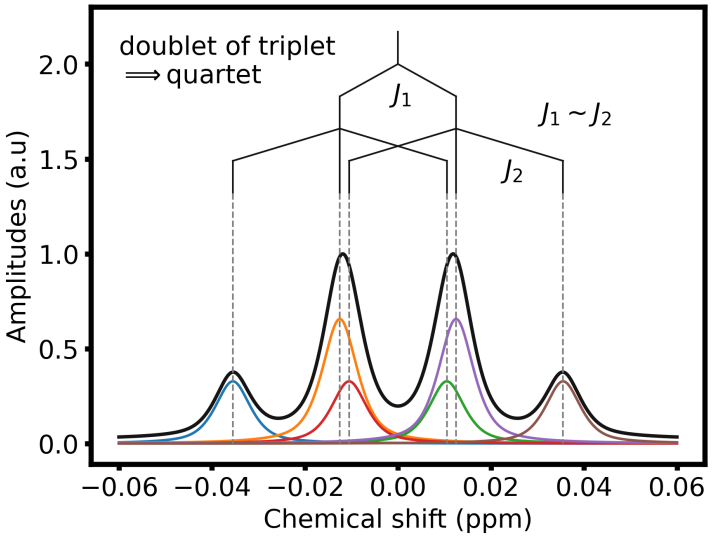


Figure 4.4: Relabeling algorithm for a doublet of triplets with nearly equal coupling constants $J_1 \sim J_2$. The 6 peaks of the doublet of triplets appear in the spectrum (lines in various colours), but their combined signal forms a multiplet (black line) resembling a quartet. As a result, the multiplet is added to the training set with its label changed from doublet of triplets to quartet.

peaks. The linewidth reduction was implemented through a smooth, sigmoid-shaped scaling function of the signal-to-noise ratio SNR. Specifically, the linewidth of each peak is multiplied by the factor

$$f(\mathrm{SNR}) = \frac{1 - f_{\min}}{1 + \exp\left(\frac{\mathrm{SNR} - \mathrm{SNR}_c}{s}\right)} + f_{\min}, \tag{4.6}$$

with $f_{\min} = 0.8$, $\mathrm{SNR}_c = 30$, and $s = 5$, so that higher SNR values result in narrower peaks, with $f(\mathrm{SNR}) \in [f_{\min}, 1)$ ensuring the linewidth is always reduced by at most a factor of $f_{\min}$ (see Figure 4.5). For low signal-to-noise ratios, the scaling factor remains close to unity, resulting in negligible linewidth modification, whereas for high signal-to-noise ratios it asymptotically approaches $f_{\min}$. This continuous parametrization avoids abrupt changes in linewidth and selectively enhances spectral resolution in regions where the signal quality supports reliable peak discrimination. The number of detected peaks was then compared to the expected count. If a mismatch was found, we evaluated whether the signal better matched another known multiplet phenotype. If so, we updated the label; if not, meaning the phenotype matched no known class, the segment was excluded from the training set.

We applied this automatic strategy to generate an additional $10'000$ synthetic samples, used to validate our frameworks.

4.3 Testing set

A critical step in the evaluation process involves assessing the model's robustness and its ability to retain predictive accuracy when confronted with real-world experimental variability, as opposed to the synthetic examples encountered during training. To perform this assessment, after the training phase, we tested the generalization performance of our

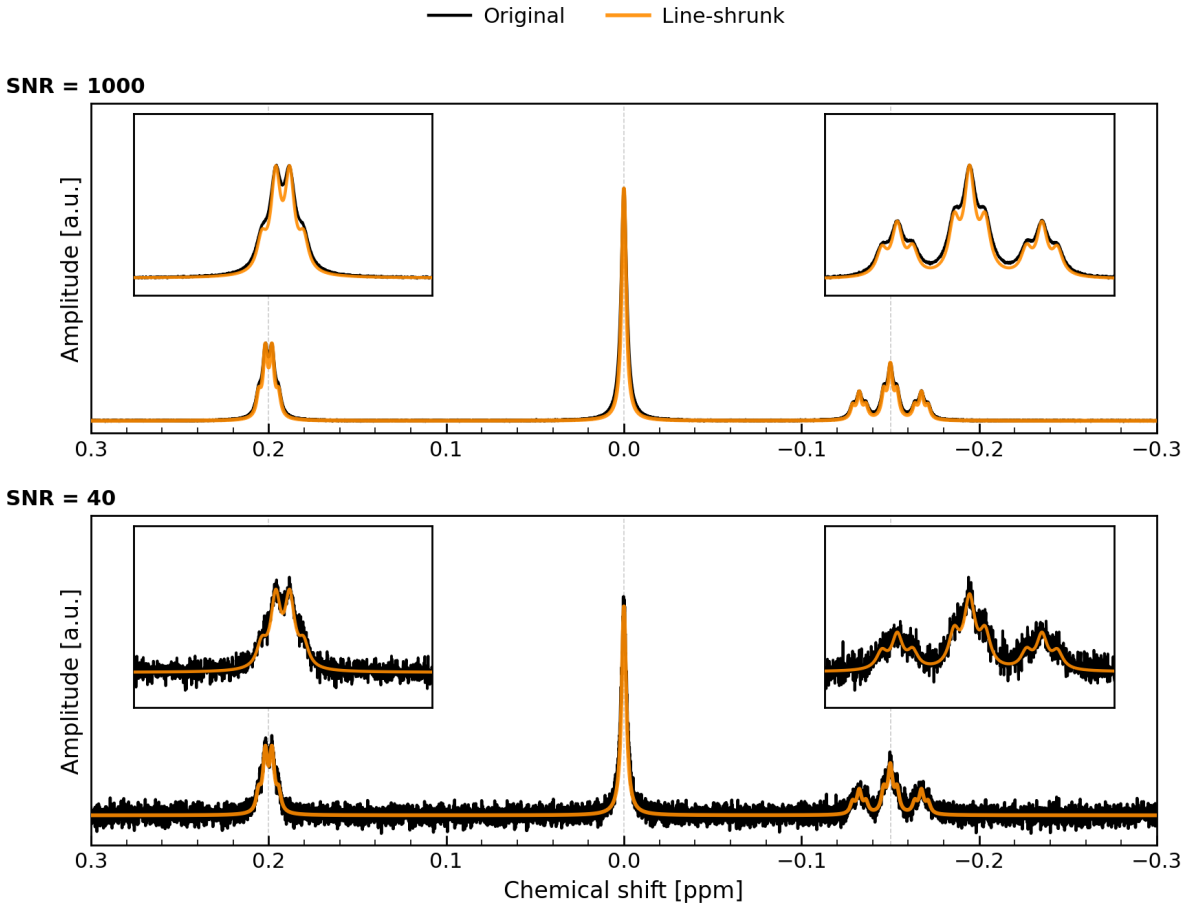


Figure 4.5: Illustration of linewidth shrinking applied to two NMR spectral segments with different signal-to-noise ratios (SNR). Each panel shows a simulated spectrum containing a singlet (0 ppm), a triplet of triplets (−0.15 ppm), and a quartet (0.20 ppm), with Lorentzian linewidths of 1.5 Hz. The original noisy spectrum is shown in black and the line-shrunk version in orange.

classification model on a dataset consisting of experimental ^{1}H NMR spectra obtained from small molecules. We define small molecules as organic compounds with molecular weights of 1000 daltons or less. This class of compounds is highly relevant in the pharmaceutical industry [88], with approximately 90% of approved therapeutic drugs falling into this category and exhibiting an average molecular weight of about 500 daltons. One-dimensional ^{1}H NMR spectroscopy serves as a powerful tool in the characterization and discovery of such compounds, offering a fast and accessible means of structural analysis. In contrast, the one-dimensional spectra of larger molecules such as proteins are characterized by significant peak overlap, and thus two-dimensional NMR techniques are typically employed for their analysis. The first prototypical implementation was evaluated on a set of ten experimental spectra (see Table 4.4), while the full probabilistic framework, MuSe Net, was subsequently tested on an expanded set of 48 experimental spectra (see Table 4.5). Both datasets were annotated by expert spectroscopists at the partner company Bruker Switzerland AG, who labelled only the peak positions within each multiplet, without delineating the multiplet boundaries.

Table 4.4: *Experimental test set.* Details of the 10 experimental spectra used to evaluate the point-estimate model [52].

Compound	Solvent	Base frequency (MHz)	Points	PPM range
Quinine	Acetone	400	65536	−3.84-16.19
Eburnamonina	CDCl3	500	65536	−1.50-14.50
Geraniol	CDCl3	400	32768	−4.10-16.45
Camphor	CDCl3	400	32768	−4.10-16.45
3-Ethyltoluol	DMSO	400	32768	−4.10-16.45
Cumol	DMSO	400	32768	−4.10-16.45
L-lactic acid	DMSO	400	32768	−4.10-16.45
Naringenin	DMSO	400	32768	−4.10-16.45
Cinchocaine	DMSO	600	32768	−4.19-16.44
Quinine	MeOD	400	32768	−4.10-16.45

Table 4.5: Composition of experimental test set used to assess the performance of the probabilistic deep learning framework MuSe Net [83].

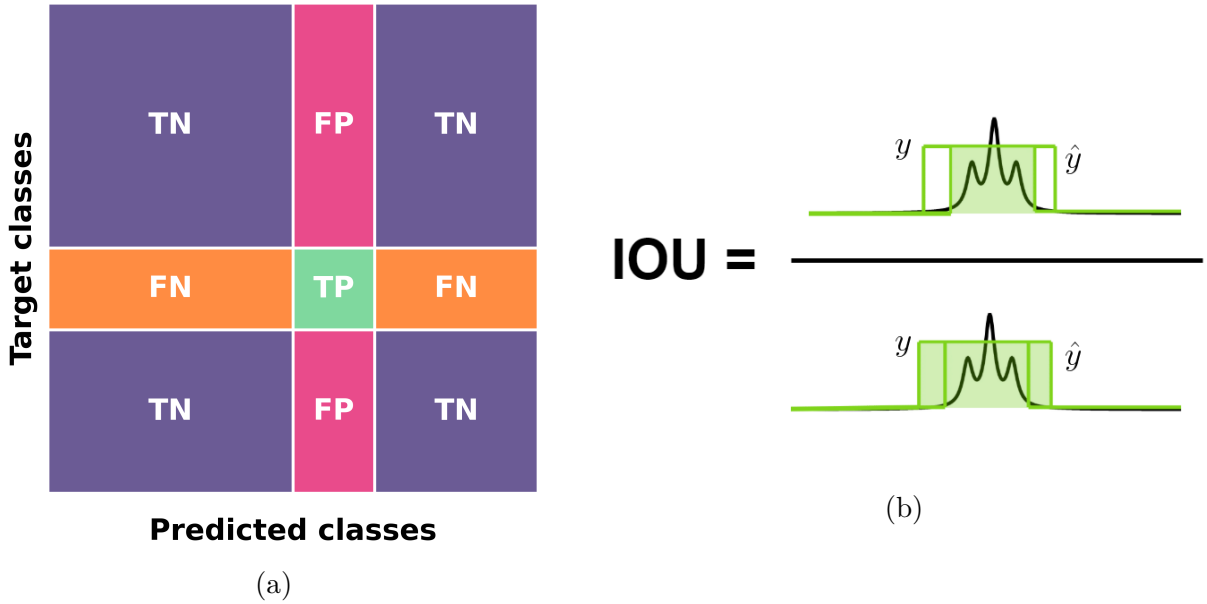
Spectrum ID	Solvent	Base frequency (MHz)	PPM range
1	DMSO	400	−2.01845, 16.00363
2	-	400	−3.49682, 16.49668
3	-	400	−3.49682, 16.49668
4	DMSO	500	−1.50134, 16.47960
5	DMSO	400	−4.10060, 16.45142
6	-	500	−1.49799, 14.49785
7	$CHCl_3$	400	−4.10034, 16.45116
8	DMSO	400	−3.82134, 16.17216

9	DMSO	800	−3.80299, 16.15380
10	DMSO	800	−3.80299, 16.15380
11	DMSO	400	−3.82134, 16.17216
12	DMSO	400	−3.82134, 16.17216
13	DMSO	800	−3.80299, 16.15380
14	DMSO	400	−4.10085, 16.45168
15	DMSO	400	−4.16632, 16.51715
16	-	600	−1.22950, 10.81841
17	-	600	−1.30918, 10.70344
18	-	600	−1.21183, 10.80079
19	-	600	−5.30835, 14.71936
20	-	400	−3.49682, 16.49668
21	DMSO	500	−4.09589, 16.44671
22	-	500	−4.09589, 16.44671
23	-	500	−4.09589, 16.44671
24	DMSO	400	−3.83777, 16.18777
25	DMSO	400	−3.83777, 16.18777
26	DMSO	400	−3.83777, 16.18777
27	DMSO	400	−3.83777, 16.18777
28	DMSO	400	−3.83777, 16.18777
29	DMSO	400	−3.82109, 16.17191
30	DMSO	400	−3.82109, 16.17191
31	DMSO	400	−3.82109, 16.17191
32	DMSO	400	−4.10060, 16.45142
33	DMSO	400	−4.10060, 16.45142
34	DMSO	400	−4.10060, 16.45142
35	$CHCl_3$	400	−4.10034, 16.45116
36	$CHCl_3$	400	−2.03029, 15.99180
37	DMSO	400	−2.01954, 16.00255
38	-	400	−2.01789, 16.00420
39	DMSO	400	−2.01813, 16.00396
40	$CHCl_3$	400	−2.02039, 16.00170
41	-	400	−3.00981, 13.00982
42	-	400	−3.00981, 13.00982
43	DMSO	600	−3.83761, 16.18843
44	DMSO	500	−1.50151, 16.47943
45	DMSO	500	−1.50125, 16.47969
46	DMSO	500	−1.50140, 16.47954
47	DMSO	500	−1.50102, 16.47992
48	-	500	−7.45280, 12.54200

4.4 Evaluation metrics for classification

Since the network generates predictions on a point-by-point basis, the most straightforward method for evaluating its performance is to assess whether each point in the spectra is correctly classified into the corresponding multiplet class. We refer to this as the pointwise approach. However, an important consideration is the evaluation of predictions not

merely at the level of individual points, but with respect to the entire multiplet structure. We refer to this second strategy as the object-wise approach. Refer to Figure 4.6 for a visual representation of these evaluation strategies.



(a) (b)

Figure 4.6: Evaluation metrics for classification. (*a*) Point-wise statistics: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) definition for a given multiplet class based on the confusion matrix entries. (*b*) Object-wise statistics: Intersection Over Union(IOU) definition in a 1D framework.

4.4.1 Point-wise approach

A common approach to evaluate performance in the multi-class classification task of distinguishing among different types of multiplet splitting patterns [89] is to compute a confusion matrix. This matrix, a square contingency table $\mathbf{C}$, compares the predicted and actual classes. The columns of the matrix represent the predicted classes, while the rows represent the actual classes. Each element C_{ij} corresponds to the number of instances belonging to class i that are predicted as class j.

From the confusion matrix, four key performance indicators can be derived: True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). These indicators characterize the model's classification errors and highlight common sources of confusion. The calculation of these indicators for each multiplet class c is illustrated below. We consider $\{x_i\}_{i=1}^{N}$ as the input data (N spectral points), and $\mathbf{f}_\theta(x)$ as the logits produced by the algorithm. The Softmax probabilities are defined as $\mathbf{p}_\theta(y|x) = \boldsymbol{\sigma}(\mathbf{f}_\theta(x))$, with the predicted labels $\hat{\mathbf{y}}_\theta = \text{argmax}_l(\mathbf{p}_\theta(y|x))$, where $\mathbf{y}$ represents the ground truth labels. For each multiplet class c, the indicators are computed as:

$$\mathrm{TP}(c) = \sum_{i=1}^{N} \delta(\hat{y}_{\theta,i} = c)\, \delta(y_i = c) \tag{4.7}$$

$$\mathrm{FP}(c) = \sum_{i=1}^{N} \sum_{k \neq c} \delta(\hat{y}_{\theta,i} = c)\, \delta(y_i = k) \tag{4.8}$$

$$\mathrm{FN}(c) = \sum_{i=1}^{N} \sum_{k \neq c} \delta(\hat{y}_{\theta,i} = k)\, \delta(y_i = c) \tag{4.9}$$

$$\mathrm{TN}(c) = N - (\mathrm{TP}(c) + \mathrm{FP}(c) + \mathrm{FN}(c)) \tag{4.10}$$

where the delta function is defined as $\delta(t) = 1$ when $t = 0$ and $\delta(t) = 0$ otherwise.

For each multiplet class, the following standard metrics can be calculated [90, 91]:

$$\mathrm{Accuracy}(c) = \frac{\mathrm{TP}(c) + \mathrm{TN}(c)}{N} \tag{4.11}$$

$$\mathrm{Precision}(c) = \frac{\mathrm{TP}(c)}{\mathrm{TP}(c) + \mathrm{FP}(c)} \tag{4.12}$$

$$\mathrm{Recall}(c) = \frac{\mathrm{TP}(c)}{\mathrm{TP}(c) + \mathrm{FN}(c)} \tag{4.13}$$

$$\text{F1 score}(c) = 2 \cdot \frac{\mathrm{P}(c) \cdot \mathrm{R}(c)}{\mathrm{P}(c) + \mathrm{R}(c)} \tag{4.14}$$

An effective classifier should aim to maximize both precision and recall, reducing false positives and false negatives. The F1 score is particularly valuable as it provides a balanced measure, combining both precision and recall into a single metric, making it especially useful when the costs of false positives and false negatives are not equivalent.

4.4.2 Object-wise approach

The Intersection Over Union (IOU) metric quantifies the overlap between the predicted and actual labels, and can be effectively employed to evaluate the alignment between the predicted signal regions and the ground truth signal regions [92]. For each class c, the IOU is computed as:

$$\mathrm{IoU}(c) = \frac{\sum_{i=1}^{N} \delta(\hat{y}_{\theta,i} = c) \cdot \delta(y_i = c)}{\sum_{i=1}^{N} \left[\delta(\hat{y}_{\theta,i} = c) + \delta(y_i = c) - \delta(\hat{y}_{\theta,i} = c) \cdot \delta(y_i = c)\right]} \tag{4.15}$$

$$\mathrm{mIoU} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \mathrm{IoU}(c) \tag{4.16}$$

where $\mathcal{C}$ denotes the set of all multiplet classes considered in the framework. The $IOU(c)$ for each class c represents the ratio of the intersection of predicted and true labels to the union of these labels, providing a measure of the prediction's accuracy for that class. The mean IOU ($mIOU$) is then calculated as the average IOU across all classes in the set $\mathcal{C}$.

Another approach to assess the model's performance in correctly locating signal regions is computing the precision-recall curve [89]. This method was originally developed

in the context of object detection in images but can be adapted to the one-dimensional case of spectral range segmentation. A label prediction Lab was defined as the connected set of points marked with a given label l, i.e., $Lab \equiv \{i \in \text{spectrum points} \mid \hat{\mathbf{y}}_i = l\}$, and surrounded either by baseline points or points marked with other labels. The confidence of each label prediction was defined as the average of the softmax outputs of the points belonging to that label prediction:

$$\text{Confidence}(Lab) = \frac{1}{n_{\text{points}}} \sum_{i \in Lab} \text{Softmax}_i(l). \tag{4.17}$$

All predicted labels, except for the baseline label, were ranked by their confidence in decreasing order. For each prediction, the IOU was computed. For each prediction in the ranked list, it was considered a true positive if its IOU was above a threshold t, typically set to 0.50 or 0.75. Otherwise, it was considered a false positive. The precision-recall curve was then computed by evaluating precision as a function of increasing recall (see Figure 4.7). Precision was measured as the ratio of true positives encountered so far to all predictions encountered so far, while recall was measured as the ratio of true positives encountered so far to the total number of predictions.

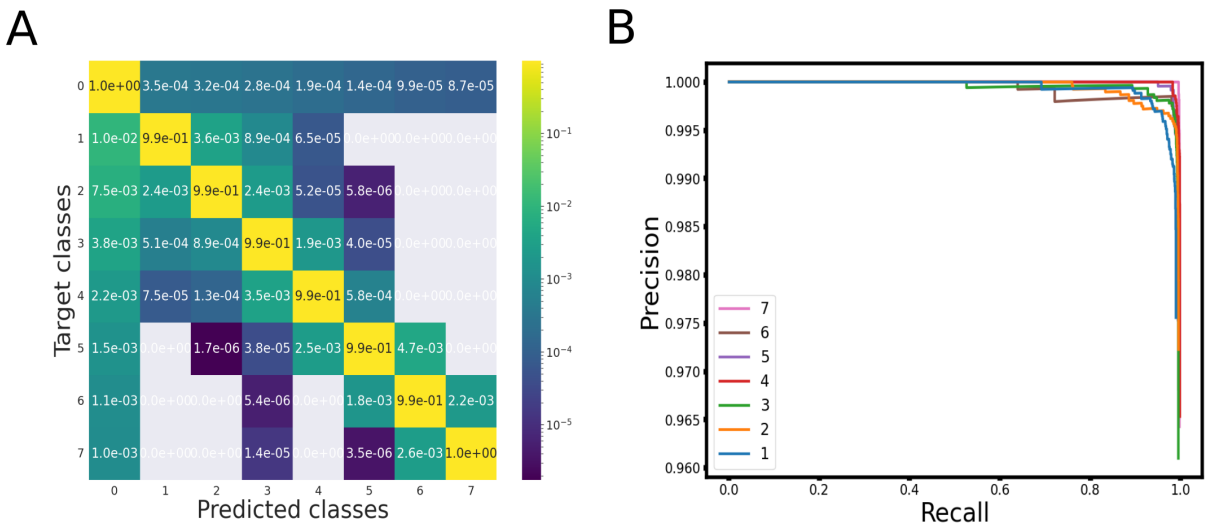


Figure 4.7: (A) Confusion matrix of the performance of the prototype network on synthetic spectra. The labels for the multiplet classes are given according to Table 4.3. (B) Object-wise approach: Precision-recall curves for each multiplet class (baseline class excluded) at threshold 75%.

The Area Under the Curve (AUC) of the precision-recall curve, also known as Average Precision (AP), was computed with an interpolation procedure over the points n of the curve:

$$\text{AP} = \sum_n (R_{n+1} - R_n) P_{\text{interp}}(R_{n+1}) \tag{4.18}$$

with

$$P_{\text{interp}}(R_{n+1}) = \max_{\tilde{R} \geq R_{n+1}} P(\tilde{R}), \tag{4.19}$$

where P and R denote the precision and recall values, respectively.

Achieving strong object-wise statistics not only ensures that a large fraction of spectral points are correctly classified, as evidenced by the point-wise statistics, but also prevents fragmented predictions. Fragmented predictions arise when different sets of spectral points within the same signal region are classified into different multiplet classes. This issue is common with algorithms that produce point-by-point outputs. When multiple class predictions exist within a signal region, each connected set is considered an individual label prediction with limited intersection with the ground truth region, which negatively impacts performance.

4.5 Evaluation of the Segmentation Framework

To assess the classification capability of our classification model, we evaluated it on a set of 10′000 synthetic spectrum segments that were generated independently of the training set. This ensured that the evaluation was not biased by any overlap with the data used during model training. The detailed performance of the classifier across all multiplet classes is summarized in the confusion matrix shown in Figure 4.7. Each row of the matrix was normalized with respect to the total number of true instances for the corresponding class, enabling a straightforward interpretation of the prediction accuracy within each ground truth category.

As seen from the matrix, the model achieves remarkably high classification accuracy, with true positive rates (diagonal elements) consistently exceeding 99% for all multiplet classes. This indicates that the model is highly reliable in correctly identifying the spectral class of each input segment.

The few misclassifications that do occur are primarily associated with the baseline class. Although these errors are rare, they generally fall into two categories: baseline points being incorrectly labeled as signal, and signal points being misclassified as baseline. Upon closer inspection of these cases, we found that the mislabeling typically occurs at the transition regions the boundaries between signal and baseline. Here, slight misalignments between the predicted and true labels often amounting to a shift of just a few points can account for the discrepancies. Such minor deviations are expected, considering the finite resolution and inherent smoothness of spectral transitions. Importantly, these small positioning mismatches do not result in a significant semantic error, such as misinterpreting pure noise as actual signal, and therefore have negligible practical impact on the downstream analysis.

From the confusion matrix, we measured accuracy, precision, recall, and F1 score. The results for each multiplet class, along with the average across all classes, are reported in Table 4.6. Accuracy is consistently above 99%, and the other classification metrics, precision, recall, and F1 score, do not fall below 98.5%. Moreover, precision and recall values exhibit minimal divergence across multiplet classes, with an average difference of just 0.21%. This indicates that the model successfully minimizes both false positives and false negatives. The balance between these two error types is further illustrated by the precision-recall curves (see Figure 4.7). An ideal classifier would maintain a precision of 1.0 across all recall values, and our model closely approaches this optimal behavior.

Table 4.6 also reports the model's object-wise performance in terms of Average Precision (AP), which corresponds to the area under the precision-recall curve, evaluated at two Intersection over Union (IoU) thresholds: 50% (AP50) and 75% (AP75). These thresholds determine the minimum overlap required for a predicted label to be considered

Table 4.6: Point-wise approach: Accuracy, precision, recall, and F1 score are reported for each multiplet class, with the final row showing the average across all classes. Object-wise approach: The Average Precision (AP), i.e., the area under the precision-recall curve, is reported at 50% and 75% thresholds for each multiplet class (excluding the baseline class). The final row presents the mean Average Precision (mAP).

	Accuracy ↑	Precision ↑	Recall ↑	F1 ↑	AP50 ↑	AP75 ↑
class 0	99.8%	99.9%	99.9%	99.9%	—	—
class 1	99.9%	98.6%	98.5%	98.6%	100%	98.9%
class 2	99.9%	99.0%	98.8%	98.9%	100%	99.5%
class 3	99.9%	99.0%	99.3%	99.1%	100%	99.5%
class 4	99.9%	99.3%	99.4%	99.3%	100%	99.8%
class 5	99.9%	99.6%	99.1%	99.4%	99.8%	99.4%
class 6	99.9%	99.1%	99.5%	99.3%	100%	99.5%
class 7	100%	99.7%	99.6%	99.7%	99.8%	99.7%
average	99.9%	99.3%	99.3%	99.3%	100%	99.5%

a true positive. As expected, increasing the IoU threshold increases the likelihood of a prediction being classified as a false positive, thereby reducing the AP values. Nevertheless, our model demonstrates remarkable robustness: the mean Average Precision (mAP) drops by only 0.47% when moving from AP50 to AP75. At AP50, all classes nearly reach perfect scores ($\geq 99.8\%$), and even at AP75, the mAP remains at an impressive 99.5%.

Among the signal classes, singlets (class 1) exhibit the most noticeable decrease in AP when raising the IoU threshold. This suggests that the localization precision for singlets is more sensitive to stricter spatial criteria, likely due to their narrower signal width compared to more complex multiplets. Nonetheless, the overall object-wise performance remains excellent, confirming that the model is effective at both fine-grained classification and precise segmentation of multiplet regions in synthetic spectra.

Training a deep neural network on synthetic data introduces a critical challenge: ensuring that the model generalizes reliably to real-world experimental data. Typically, supervised learning models operate under the assumption that the input samples in the training set are drawn from the same underlying probability distribution as the samples encountered during testing. This assumption can be expressed as $P_{\text{train}}(\mathbf{y}|\mathbf{x}) = P_{\text{test}}(\mathbf{y}|\mathbf{x})$, where $\mathbf{x}$ represents the input feature vector, and $\mathbf{y}$ is the corresponding label vector. However, when applying these models to experimental spectra, the distribution of the test data may differ significantly from that of the synthetic training data, leading to a discrepancy: $P_{\text{train}}(\mathbf{y}|\mathbf{x}) \neq P_{\text{test}}(\mathbf{y}|\mathbf{x})$. This deviation, known as a distribution shift [93], can adversely affect the model's performance and lead to a reduction in prediction accuracy.

Recent advancements in deep learning have demonstrated that it is possible to achieve good results on experimental NMR data even when the models are trained on synthetic datasets. Notable applications include spectral denoising [94], spectral reconstruction of Non-Uniformly Sampled (NUS) datasets [95, 96], peak picking, and spectral deconvolution [46, 49]. These studies show that, despite the challenges posed by distribution shifts, neural networks can be trained on synthetic data and still generalize effectively to real-world experimental spectra.

To assess the generalization capabilities of our classification model, we tested it on a set of experimental ^{1}H NMR spectra from 10 small molecules—organic compounds with a

molecular weight of 1000 daltons or less. Each spectrum contains either 32′768 or 65′536 data points (see Table S3 for more details). The model's predictions were displayed by associating each multiplet class with a different color and shading the corresponding regions of the spectrum with the predicted class color (see Figure 4.8).

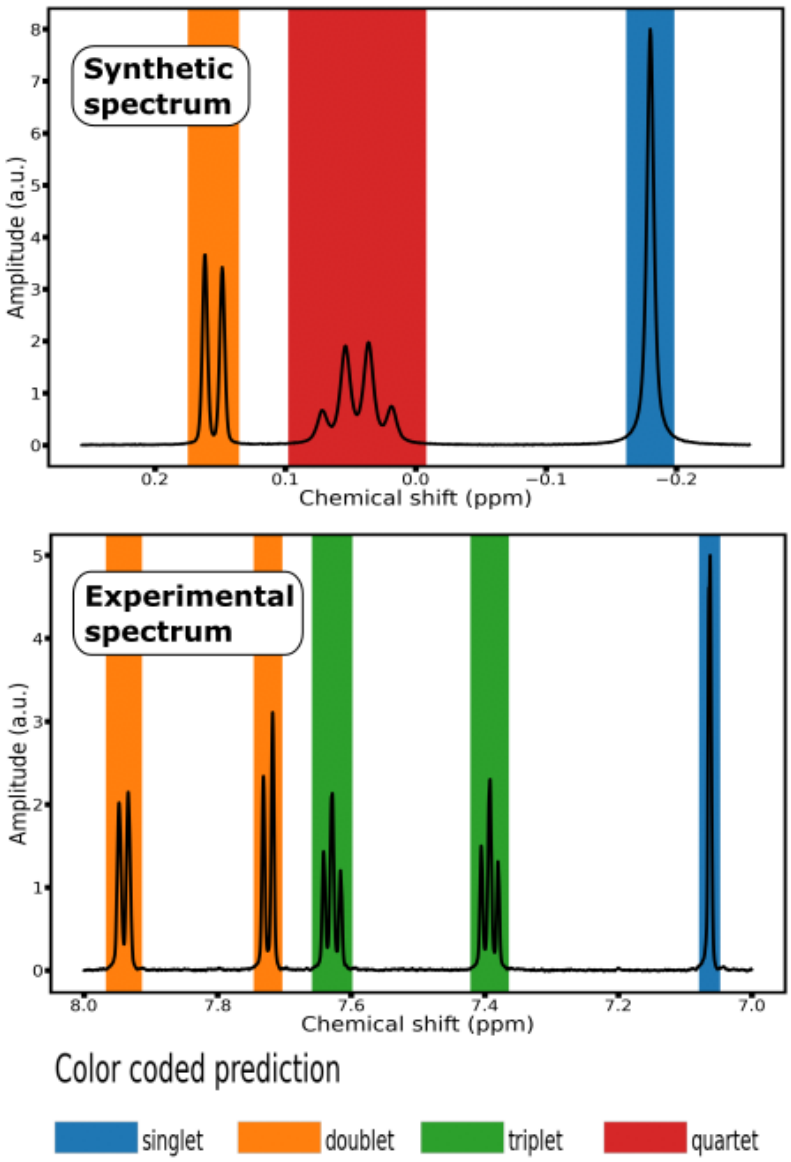


Figure 4.8: Color coded prediction of 1H NMR Spectra: (top) segment of a synthetic spectrum; (bottom) segment of an experimental spectrum from the testing set of 10 small molecules.

The experimental testing set was annotated by NMR experts, who labeled each peak with key features: position in ppm, amplitude, line width, line shape, and the multiplet class it belongs to. This annotation served as the ground truth for evaluating the model's classification accuracy.

For a quantitative assessment of the model's performance on the experimental data, we created a confusion matrix (see Figure 4.9). Each annotated peak was checked to see if its position fell within a spectral region labeled with the correct multiplet class.

The model performed well, accurately classifying all doublets, sextets, and septets in the testing set. The true positive rates for the remaining classes were consistently above 80%, showing that the model can generalize effectively despite the distribution shift between synthetic and experimental data.

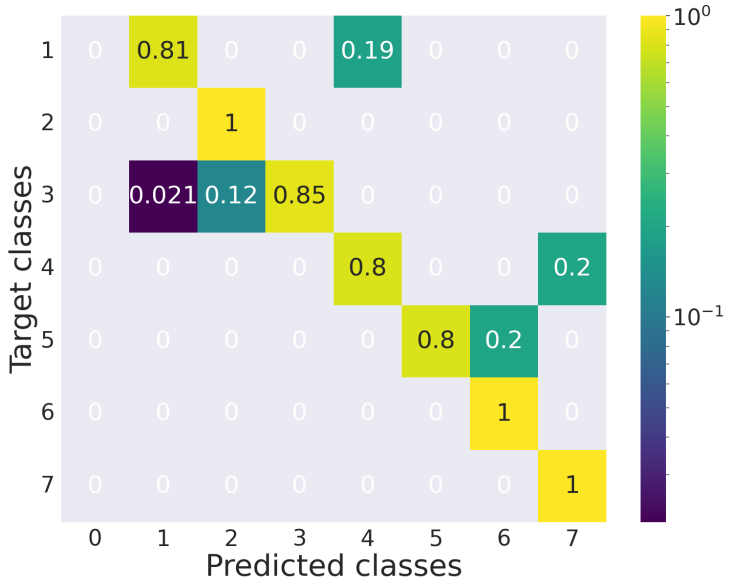


Figure 4.9: Confusion matrix of the performance on the 10 ^{1}H NMR spectra of the testing set. Note that the baseline class appears as a predicted class (column) but not as a target class (row), since the annotations provided by Bruker Switzerland AG label only the peak positions of multiplets, leaving the baseline regions without ground truth labels.

Chapter 5

Identifying multiplet splitting patterns

5.1 Uncertainty quantification and Out Of Distribution detection

In all quantitative sciences, it is essential that results are accompanied by uncertainty estimates, as they provide critical information about the reliability of the findings. However, standard deep neural networks (DNNs) typically struggle to offer meaningful measures of predictive uncertainty. A key reason for this limitation lies in the behavior of the Softmax activation function, which tends to produce probability outputs that are overly confident—even when the model's prediction is incorrect [97]. This issue becomes particularly evident when the network is presented with inputs that deviate from the training distribution. In our task of detecting and classifying multiplets, these deviating inputs typically include previously unseen splitting patterns or overlapping multiplets. In such cases, the model may assign a high probability to an erroneous class, giving a false impression of confidence. This overconfidence undermines the trustworthiness of the model, casting doubt even on predictions that happen to be correct.

To ensure that deep learning systems yield reliable and interpretable results, especially in scientific contexts, it is crucial that they incorporate mechanisms for uncertainty estimation. Ideally, a well-designed model should not only recognize when it is uncertain, but should also express this uncertainty in a meaningful way—that is, by failing gracefully. Furthermore, for a model to be considered calibrated, its predicted confidence levels should correspond closely to the actual likelihood of correctness: predictions with high confidence should be correct more often than those with low confidence.

Machine learning algorithms are subject to various sources of uncertainty, which can generally be categorized into two main types: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty is associated with the inherent stochastic variability between the inputs and outputs of the model, making it irreducible. This type of uncertainty is particularly prominent in regions of the feature space where there is significant class overlap, meaning that different classes share similar features and therefore cannot be easily distinguished. On the other hand, epistemic uncertainty arises from the model's limited knowledge about the optimal predictive function. This uncertainty can be reduced or even eliminated as more training data is provided, ultimately approaching zero in the limit of infinite data.

In the context of machine learning, uncertainty estimation can be extremely valuable, especially when the model encounters data that it has not been trained on. Specifically,

epistemic uncertainty tends to increase when the algorithm is exposed to inputs whose features deviate from those seen during training. These instances are typically considered Out Of Distribution (OOD) data, as they lie outside the feature space of the training set. Uncertainty estimation allows the model to recognize when it encounters such OOD samples, which can then be flagged for additional analysis or sent to a human expert for further review. This capability is referred to as Out Of Distribution detection (OOD detection). In this framework, the classes represented in the training set are referred to as In Distribution (ID) classes, while everything that deviates from this known space is labeled as Out Of Distribution (OOD).

In our multiplet classification algorithm, the main sources of OOD instances were overlapping multiplets and multiplets exhibiting second-order splitting patterns [98], neither of which were included in the training set. By incorporating uncertainty quantification and OOD detection, we were able to enhance our classification system by introducing a general multiplet class to account for these OOD multiplets, ensuring that the algorithm could handle such cases more effectively.

To obtain a reliable uncertainty estimate, a fundamental shift in the framework was required. Specifically, we transitioned from a Point-Estimate algorithm, which produces a single deterministic output for each input, to a Probabilistic approach capable of generating a predictive posterior distribution over the possible outputs. In this new framework, both the prediction and its associated uncertainty are derived from the first two moments of the distribution—the mean and the variance. This allows for a more nuanced understanding of the model's predictions, as it captures not only the expected output but also the uncertainty surrounding it.

The Bayesian approach provides the theoretical foundation for this type of uncertainty-aware machine learning. In this framework, a prior distribution on the model parameters is updated based on the observed data to infer a posterior distribution over the possible outcomes. This approach, known for its robustness in uncertainty estimation, underpins the most widely used models for such tasks, Bayesian Neural Networks (BNNs). BNNs treat the model parameters as random variables and, through Bayesian inference, provide a full distribution over possible predictions rather than a single point estimate [99]. Despite their principled nature, the practical application of BNNs is often limited by the analytically intractable nature of their posterior distributions.

To address this challenge, several inference techniques have been developed, including Markov Chain Monte Carlo (MCMC) methods and Variational Inference [100, 99]. These methods approximate the posterior distribution, making BNNs more computationally feasible. However, the performance of BNNs is still highly dependent on the choice of the prior distribution, which introduces another layer of complexity. In addition to traditional BNNs, alternative approaches have been proposed to approximate the predictive posterior distribution more efficiently. These methods often aim to reduce computational overhead while maintaining a more straightforward implementation, making them appealing for practical applications where computational resources are limited.

For our automatic multiplet classification task, we evaluated three state-of-the-art deep learning frameworks for uncertainty quantification: Spectral Normalized Gaussian Process (SNGP) [101, 102], Monte Carlo Dropout (MCD) [103], and Deep Ensembles (ENS) [104]. Each of these approaches offers a unique method for incorporating uncertainty into the model's predictions. SNGP, in particular, stands out by providing both the classification of splitting patterns and a corresponding covariance matrix, which quantifies the uncertainty in the predictions. On the other hand, MCD and ENS esti-

mate uncertainty by aggregating the outputs of multiple instances of the network model. Specifically, MCD achieves this by applying dropout at both training and inference stages, while ENS creates an ensemble of independently trained models, where uncertainty is captured through the variation in predictions across the ensemble.

All three algorithms were integrated into our system by starting with the base architecture discussed in Section 4.1.1. In the following paragraphs we illustrate the implementation of these uncertainty quantification methods which involved minimal modifications to the original model.

5.1.1 Spectral Normalized Gaussian Process

Spectral,normalized Gaussian Process (SNGP) [101, 102] is a recently proposed deterministic deep learning framework designed to provide reliable uncertainty estimates using a single network, thereby eliminating the computational and architectural complexity typically associated with Bayesian methods. Rather than relying on multiple models to capture predictive variability, SNGP introduces structural modifications to the network that enable it to quantify uncertainty in a principled and efficient manner.

The central idea underlying SNGP is the notion that a robust and trustworthy deep learning model should possess the ability to infer whether a test input $\mathbf{x}$ belongs to the in,domain data distribution $p(\mathbf{x} \in \text{ID})$. This capability hinges on the model's sensitivity to the distance between the test input $\mathbf{x}$ and the training samples in the input space. Crucially, this notion of distance must be preserved throughout the feature representations learned by the hidden layers and must be interpretable by the output layer to support effective uncertainty estimation. Mathematically, the hidden mapping h must satisfy a bi,Lipschitz condition to be distance,preserving, bounded by constants L_1 and L_2:

$$L_1 d(x_1, x_2) \leq d(h(x_1), h(x_2)) \leq L_2 d(x_1, x_2) \tag{5.1}$$

To enforce this behavior, SNGP employs spectral normalization on the weights of the hidden dense layers. This regularization technique ensures Lipschitz continuity, thereby preserving relative distances in the feature space. It does so by normalizing the weight matrix W_l at each layer using its estimated spectral norm $\hat{\lambda}$ via the following update rule:

$$W_l = \frac{cW_l}{\hat{\lambda}} \quad \text{if } c < \hat{\lambda} \tag{5.2}$$

where c is a bounding hyperparameter. At the output level, the standard dense layer is replaced by a Gaussian process (GP), which models the predictive distribution over the output classes. The GP is constructed with a prior multivariate normal distribution whose covariance depends explicitly on the feature distances, capturing the local geometry of the learned representations. To maintain scalability, SNGP approximates this GP prior using Random Fourier Features Φ.

During inference, the model computes the posterior distribution of the GP given the training data. This yields both the maximum a posteriori (MAP) estimate for classification and a full covariance matrix encoding predictive uncertainty. Specifically, using a Laplace approximation, the posterior precision matrix (inverse covariance) for class k is efficiently updated during training as:

$$\hat{\Sigma}_k^{,1} = I + \sum_{i=1}^{N} \hat{p}_{i,k}(1, \hat{p}_{i,k}) \Phi_i \Phi_i^\top \tag{5.3}$$

To make the uncertainty estimation tractable in practice, the diagonal elements of the covariance matrix are extracted and used to quantify the model's confidence in its predictions.

Crucially, the specific kind of uncertainty extracted by SNGP is an input distance,aware uncertainty. Because the GP output layer utilizes a radial basis function kernel, the posterior predictive variance for a given input is directly characterized by its spatial distance from the training data in the hidden representation space. As a test example moves further away from the observed training manifold, this predictive variance monotonically increases. Consequently, for highly uncertain out,of,domain inputs, the model natively backs off to a discrete uniform, or maximum entropy, predictive distribution. This explicit reflection of spatial uncertainty makes SNGP especially valuable in safety,critical applications or scenarios with significant distributional shift.

5.1.2 Monte Carlo Dropout

Monte Carlo Dropout (MCD), introduced by Gal and Ghahramani [103], established a theoretical link between standard deep neural networks with dropout and probabilistic deep Gaussian processes. Specifically, they demonstrated that applying dropout before every weight layer in a neural network enables an efficient approximation of a deep Gaussian process, thereby transforming a traditionally deterministic model into a stochastic one capable of modeling predictive uncertainty. This equivalence, combined with the minimal computational overhead required to implement dropout, has made MCD a widely adopted and effective method for uncertainty estimation in a variety of classification tasks, including applications in dermatology [105], image recognition [106], and data-efficient learning [107].

In our implementation, we integrated dropout layers with a dropout probability of $p = 0.5$ before each hidden dense layer in the network architecture. Importantly, we configured the dropout to remain active during both training and inference phases, contrary to its traditional use where dropout is disabled at test time. This enables the model to produce a distribution over outputs by randomly dropping a subset of neurons during each forward pass, effectively turning a single neural network into an ensemble of subnetworks sampled from the approximate posterior.

During inference, we performed $M = 100$ stochastic forward passes through the network to obtain an empirical approximation of the predictive distribution $p_{\theta_m}(\hat{y}|x)$, where θ_m denotes the randomly perturbed parameters in the m-th realization of the network, x is the input sample, and $\hat{y}$ the output label. The final classification prediction was then computed as the mean over all sampled predictive distributions:

$$\bar{p}(\hat{y}|x) := \mathbb{E}_{p(\theta|\mathcal{D})}[p_\theta(\hat{y}|x)] = \frac{1}{M}\sum_{m=1}^{M} p_{\theta_m}(\hat{y}|x), \tag{5.4}$$

where $p(\theta|\mathcal{D})$ represents the (approximate) posterior distribution over the model parameters given the training dataset $\mathcal{D} = \{x^{\text{train}}, y^{\text{train}}\}$.

To quantify the model's predictive uncertainty, we decomposed the total *predictive uncertainty (PU)* into two complementary components: *aleatoric uncertainty (AU)*, which captures the intrinsic noise in the data, and *epistemic uncertainty (EU)*, which reflects uncertainty in the model parameters due to limited data. The total predictive uncertainty

was computed as the normalized entropy of the mean predictive distribution:

$$PU \equiv H_1 := H[\bar{p}(\hat{y}|x)] = -\frac{1}{\log|\mathcal{C}|}\sum_{c\in\mathcal{C}}\bar{p}(\hat{y}=c|x)\log\bar{p}(\hat{y}=c|x), \tag{5.5}$$

where $\mathcal{C}$ denotes the set of target classes considered in our classification task.

To isolate the aleatoric component, we computed the expected entropy of the individual stochastic forward passes:

$$\begin{aligned} AU \equiv H_2 &:= \mathbb{E}_{p(\theta|\mathcal{D})}[H[p_\theta(\hat{y}|x)]] \\ &= -\frac{1}{M\log|\mathcal{C}|}\sum_{m=1}^{M}\sum_{c\in\mathcal{C}} p_{\theta_m}(\hat{y}=c|x)\log p_{\theta_m}(\hat{y}=c|x), \end{aligned} \tag{5.6}$$

This quantity captures the variability in predictions due to class overlap and label ambiguity inherent in the input data.

The epistemic uncertainty, finally, was estimated as the *mutual information (MI)* between predictions and model parameters:

$$EU = PU - AU \iff MI := H_1 - H_2, \tag{5.7}$$

which measures the reduction in uncertainty about the output gained from knowing the specific parameters of the model. In practice, high epistemic uncertainty is indicative of inputs lying far from the training distribution or in regions with sparse training data, thus providing a valuable signal for out-of-distribution detection and model calibration.

5.1.3 Deep Ensembles

A *deep ensemble* [104] is a collection of independently trained deep neural networks whose predictions are aggregated to enhance both predictive performance and robustness. This approach has consistently shown superior results compared to single-network models, with its effectiveness supported by both theoretical insights and empirical success across a wide range of tasks [108].

In our framework, we constructed an ensemble by training $M_s = 10$ instances of the primary network architecture. Each instance was initialized independently using weights sampled from a shared prior distribution $p(\theta)$, ensuring diversity across the ensemble members. During inference, all M_s networks were evaluated on the same test set, and their predictive outputs were collected.

The final classification result was obtained by averaging the predictive distributions across the ensemble:

$$\bar{p}(\hat{y}|x) := \frac{1}{M_s}\sum_{m=1}^{M_s} p_{\theta_m}(\hat{y}|x), \tag{5.8}$$

where θ_m denotes the parameter set of the m-th network in the ensemble.

To quantify the model's epistemic uncertainty, we followed the same procedure used for Monte Carlo Dropout (see Equations 5.4–5.7). Specifically, we computed the normalized entropy of the average predictive distribution to obtain the total predictive uncertainty, estimated the expected entropy of the individual ensemble members for aleatoric uncertainty, and derived the epistemic uncertainty from their difference via the mutual information:

$$EU = H_1 - H_2. \tag{5.9}$$

This method benefits from the inherent diversity of the independently trained networks, which captures the variability in the learned representations and provides a reliable estimate of epistemic uncertainty, particularly in regions far from the training distribution.

5.1.4 Calibration

To ensure that the predictive probabilities produced by our deep learning model are not only accurate but also properly reflect the true likelihood of being correct, we conducted a thorough evaluation of model calibration. Calibration assesses the degree to which a model's predicted confidence aligns with actual observed accuracy. To this end, we utilized reliability diagrams, which plot the model's accuracy against its confidence scores—ideally, a perfectly calibrated model should yield points lying on the diagonal, where confidence matches accuracy. For post-hoc calibration, we adopted temperature scaling [97], a simple yet effective method that adjusts the model's softmax outputs to better reflect calibrated probabilities (see Figure 5.1).

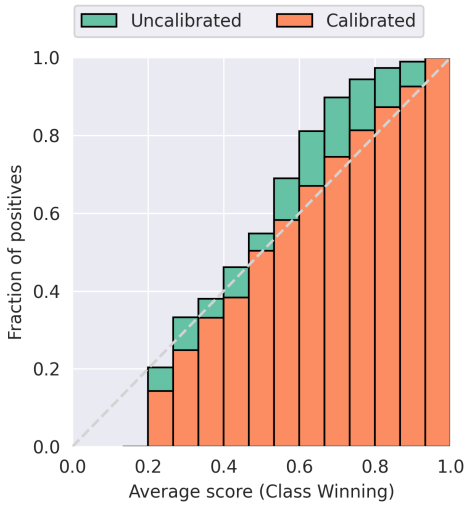


Figure 5.1: Reliability diagram of the MuSe Net framework using deep ensembles on the synthetic validation set, shown before (green) and after (orange) applying temperature scaling for calibration.

In the case of ensemble-based uncertainty estimation methods such as Monte Carlo Dropout and Deep Ensembles, we first aggregated the outputs from multiple stochastic forward passes or ensemble members, respectively. Calibration was then applied to the resulting predictive distributions to improve reliability [109].

To quantitatively assess calibration quality, we computed multiple metrics, including Negative Log-Likelihood (NLL), Brier Score (BS), and the class-wise Expected Calibration Error (cwECE) [110], which together provide a comprehensive evaluation of both

confidence calibration and class-specific reliability:

$$\text{NLL} = -\sum_{i=1}^{N} y_i \log\left(p_{\theta,\,i}\right) \tag{5.10}$$

$$\text{BS} = \frac{1}{N}\sum_{i=i}^{N}(y_i - p_{\theta,\,i})^2 \tag{5.11}$$

$$\text{cwECE} = \frac{1}{|C|}\sum_{c=0}^{|C|}\text{class-c ECE} \tag{5.12}$$

$$\text{class-c ECE} = \sum_{m=1}^{M}\frac{|B_m^{(c)}|}{N}|\text{acc}(B_m^{(c)}) - \text{conf}(B_m^{(c)})| \tag{5.13}$$

$$\text{acc}(B_m^{(c)}) = \frac{1}{|B_m^{(c)}|}\sum_{i\in B_m^{(c)}}\delta(\hat{y}_{\theta,i} = c)\,\delta(y_i = c) \tag{5.14}$$

$$\text{conf}(B_m^{(c)}) = \frac{1}{|B_m^{(c)}|}\sum_{i\in B_m^{(c)}} p_{\theta,i}^{(c)} \tag{5.15}$$

where $B_m^{(c)}$ is the set of spectral points belonging to the multiplet class c in the $m-$bin.

The Negative Log-Likelihood (NLL) serves as a principled measure for evaluating probabilistic predictions, particularly in classification tasks. It is minimized when the predicted probabilities align perfectly with the true conditional distribution $\mathbf{\hat{p}(y|x)}$ [111], effectively rewarding models that assign high confidence to correct predictions and penalizing those that are both incorrect and overconfident. Due to its logarithmic form, NLL is especially sensitive to confidently wrong predictions, making it a rigorous test of calibration quality.

The Brier Score (BS) provides an alternative perspective by measuring the mean squared error between the predicted probabilities and the actual outcomes. Unlike NLL, the Brier Score applies a quadratic penalty, which is bounded and symmetric, thus offering a more interpretable assessment of the distance between predicted and true probabilities. It is particularly useful when the goal is to understand average deviations in confidence, regardless of whether the model is over- or under-confident.

Finally, to summarize the results from the reliability diagram in a single scalar quantity, we report the Expected Calibration Error (ECE) and its class-wise extension, cwECE. ECE provides a global view of miscalibration by averaging the absolute difference between confidence and accuracy across M confidence bins:

$$\text{ECE} = \sum_{m=1}^{M}\frac{|B_m|}{n}\,|\text{acc}(B_m) - \text{conf}(B_m)| \tag{5.16}$$

where B_m is the set of samples in bin m, n is the total number of samples, and $\text{acc}(B_m)$ and $\text{conf}(B_m)$ are the accuracy and mean confidence within that bin. cwECE extends this by averaging the calibration error over individual classes [110]:

$$\text{cwECE} = \frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{M}\frac{|B_m^k|}{n_k}\left|\text{acc}_k(B_m^k) - \text{conf}_k(B_m^k)\right| \tag{5.17}$$

where K is the number of classes and n_k is the number of samples belonging to class k. This distinction is especially important in multiclass or imbalanced settings, where certain classes may be significantly more miscalibrated than others. The use of cwECE

in our analysis thus enables a more fine-grained understanding of the model's calibration behavior across the full set of predicted categories.

5.2 MuSe Net results

5.2.1 Model evaluation and calibration

To obtain high-quality multiplet segmentation with reliable and well-characterised predictive uncertainty, we conducted a comprehensive and systematic evaluation of three probabilistic deep learning strategies: spectral normalised Gaussian processes (SNGP), Monte Carlo dropout (MCD), and deep ensembles (ENS). These methods were selected for their complementary theoretical foundations and widespread applicability in uncertainty-aware modelling. Each model architecture was trained independently on the same dataset, ensuring that performance differences could be attributed solely to the uncertainty quantification approach rather than confounding factors such as data composition or preprocessing. To facilitate a robust and unbiased comparison under controlled conditions, we constructed a large, hold-out benchmark set comprising 10′000 synthetic samples, which were generated in a separate run using the same automated procedure employed for the training set. This design preserved statistical independence while maintaining strict distributional consistency, thereby allowing us to evaluate generalisation and calibration under i.i.d. conditions that mirror the training regime but exclude any data leakage.

Our validation framework was designed with dual emphasis: to measure not only the raw classification accuracy of the predicted multiplet segments, but also the fidelity and interpretability of the model's associated confidence scores. This is especially critical in scientific and analytical applications where downstream decision-making often hinges on the quality of probabilistic output, such as threshold-based inference, active learning, or uncertainty-aware hypothesis testing. For evaluating classification performance, we employed three complementary metrics: balanced accuracy, which mitigates the effect of class imbalance by averaging recall across all categories; the F1 score, which captures the harmonic mean of precision and recall and penalizes both false positives and false negatives; and the mean Intersection over Union (mIOU), which assesses the degree of overlap between predicted and ground-truth regions, thus quantifying structural fidelity beyond pointwise correctness. These metrics jointly provide a comprehensive picture of segmentation accuracy, capturing both fine-grained labelling behaviour and broader region-level consistency.

However, predictive performance alone is not sufficient in contexts where uncertainty quantification is essential. We therefore evaluated the calibration of predicted probabilities to assess how well the model's confidence estimates reflect actual empirical correctness. Ideally, a well-calibrated model should produce predictive distributions such that, for any given confidence level (e.g., 90%), the observed accuracy for predictions made at that level is also approximately 90%. To diagnose and quantify calibration, we utilised both graphical and scalar methods. Reliability diagrams provided a visual summary of calibration behaviour by plotting empirical accuracy against predicted confidence across binned intervals, where perfect calibration corresponds to alignment with the identity line.

Recognising the tendency of deep neural networks to produce overconfident outputs, we applied a post-hoc recalibration technique known as temperature scaling [97] to all

three methods. This approach introduces a scalar temperature parameter that rescales the logit outputs prior to the softmax transformation, with the temperature optimised to minimise NLL on a held-out calibration set. Notably, temperature scaling preserves the relative ranking of class logits, and hence does not affect the final class predictions, only the sharpness of the predicted probability distribution. It is widely regarded as a simple yet effective method for correcting calibration errors without modifying the model architecture or training procedure. An illustrative example of its impact on prediction confidence and reliability is shown in Figure 5.1.

Following calibration, we compared the overall performance of the three approaches, considering both segmentation accuracy and uncertainty quality. As summarised in Table 5.1, the deep ensemble (ENS) method demonstrated the most favourable trade-off, achieving consistently high classification accuracy while maintaining well-calibrated confidence estimates across all metrics. This finding aligns with existing literature that identifies deep ensembles as one of the most effective and robust strategies for uncertainty quantification in neural networks [101, 102]. Although Monte Carlo dropout (MCD) achieved comparable results in terms of predictive performance, it required approximately ten times more forward passes at inference time, resulting in significantly higher computational overhead and reduced scalability. In contrast, the deep ensemble approach provided comparable or better accuracy with substantially lower runtime complexity per unit prediction.

Based on these observations, we adopted a deep ensemble of 10 independently trained, identically structured neural networks as the uncertainty estimation backbone of the MuSe Net framework. Each model in the ensemble was initialised and trained independently, allowing the ensemble to capture a diverse set of learned representations and thereby better approximate the posterior distribution over model predictions. Averaging predictions across ensemble members enhanced overall accuracy and robustness, while the variability among outputs enabled meaningful estimation of predictive uncertainty. This design supports a principled Monte Carlo approximation of the posterior predictive distribution and integrates seamlessly into downstream uncertainty-aware applications.

Finally, quantitative analysis conducted on the synthetic validation data confirmed the effectiveness of our model selection. As shown in Figure 5.2, MuSe Net achieved high segmentation performance across all multiplet splitting patterns represented in the training set, with both false positives and false negatives minimised to near-optimal levels. These results underscore the framework's ability to generalise effectively across a range of chemically relevant signal structures and validate the suitability of our deep ensemble approach for uncertainty-aware NMR signal interpretation. Taken together, this extended validation effort, spanning model comparison, probabilistic calibration, post-hoc recalibration, and rigorous quantitative benchmarking, supports the deployment of a deep ensemble configuration as a reliable and well-calibrated backbone for multiplet segmentation within the MuSe Net framework.

5.2.2 Generalisation on experimental spectra

To determine the extent to which MuSe Net can be reliably and routinely integrated into real-world NMR analysis pipelines, we evaluated its performance on an experimental test set comprising 48 ^{1}H NMR spectra of small organic molecules. Robust generalisation to unseen, real-world data is a central benchmark for the practical utility of any deep learning system. In this context, training on synthetic data—while highly beneficial in terms

Table 5.1: Comparison of models in terms of classification performance and calibration. Acc refers to balanced accuracy, F1 to the F1 score, mIOU to the mean class-wise intersection over union, NLL to negative log likelihood, BS to the Brier score, and cwECE to the class-wise expected calibration error.

Model	Acc ↑	F1 ↑	mIOU ↑	NLL ↓	BS ↓	cwECE ↓
SNGP	97.5%	95.6%	91.7%	0.099	$2.4 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$
MCD	98.8%	98.0%	96.1%	0.028	$1.2 \cdot 10^{-3}$	$7.3 \cdot 10^{-3}$
ENS	**99.0%**	**98.2%**	**96.4%**	**0.026**	$\mathbf{1.1 \cdot 10^{-3}}$	$\mathbf{4.9 \cdot 10^{-4}}$

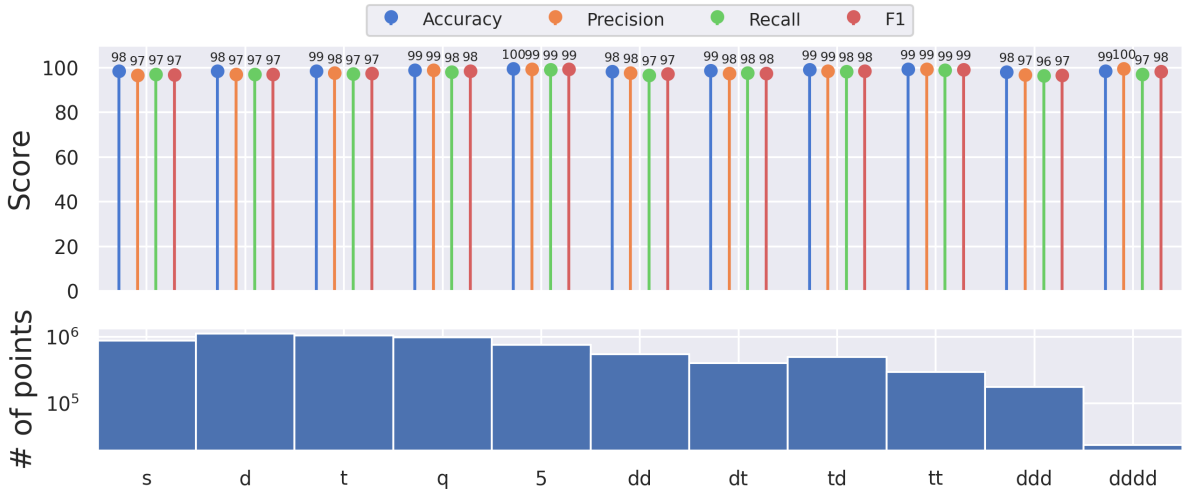


Figure 5.2: Evaluation metrics for point-by-point MuSe Net classification on the synthetic validation set (top panel), along with the ground truth distribution of spectral points among multiplet classes (bottom panel). Here, *s* stands for singlets, *d* for doublets, *t* for triplets, *q* for quartets, 5 for quintets, *dd* for doublets of doublets, *dt* for doublets of triplets, *td* for triplets of doublets, *tt* for triplets of triplets, *ddd* for doublets of doublets of doublets, and *dddd* for doublets of doublets of doublets of doublets. Scores are given in percentage.

of data availability, controlled variability, and ground-truth annotation—inevitably introduces a distributional gap between synthetic and experimental domains. Despite the inclusion of various realistic perturbations in our synthetic training set, such as small phase and baseline distortions, varying levels of noise, rooftop effects, and diverse lineshapes, we expect that the resulting distribution of synthetic features, denoted by $P_{\text{synt}}(\mathbf{x}_{\text{synt}})$, differs from the corresponding distribution observed in real experimental spectra, denoted by $P_{\text{exp}}(\mathbf{x}_{\text{exp}})$. This discrepancy, commonly referred to as covariate shift [112], can degrade model performance if not accounted for or empirically evaluated.

Figure 5.3 presents the classification performance of MuSe Net across the experimental dataset, summarised using recall as the primary evaluation metric. Excluding the classes `dt` (doublet of triplets) and `ddd` (doublet of doublet of doublets), MuSe Net consistently achieves recall rates above 83%, reaching up to 100% for `q` (quartet) patterns. Upon visual inspection of the misclassified cases, most `dt` errors appear to arise from difficulty in resolving the component peaks, which frequently merge due to insufficient resolution or overlapping coupling constants (see bottom row of Figure 5.4). Notably, even in such ambiguous scenarios, MuSe Net often detects at least one of the correct splittings, typically the doublet, indicating a partial but meaningful recognition of the signal's multiplet structure.

Multiplets with a higher number of coupling constants naturally exhibit greater structural complexity, making them harder to classify reliably. In particular, certain parameter combinations—such as similar coupling constants or broadened lines—can cause distinct multiplet types to appear visually similar, leading to confusion between phenotypically adjacent classes. As discussed previously, this phenomenon is reflected in cases where, for example, a `ddd` with one dominant coupling constant is misclassified as two overlapping `dd` patterns. Similarly, a `ddd` consisting of four merged central peaks flanked by two additional peaks may be mistaken for a `td` (triplet of doublets), highlighting the limitations of static phenotype boundaries in the presence of significant peak overlap (see Figure 5.5, bottom row). Despite these challenges, MuSe Net demonstrated considerable robustness in classifying diverse manifestations of the `ddd` class, as evidenced by successful predictions in more structurally resolved examples (top row of Figure 5.5).

Regarding the `dddd` class (doublet of doublet of doublets of doublets), the experimental dataset contained only a single annotated example. MuSe Net produced a partially fragmented prediction: one subset of the signal was correctly labelled as `dddd`, while another was misclassified as `ddd`. While not fully accurate, this outcome still suggests sensitivity to the underlying structure. Nonetheless, drawing general conclusions from a single observation is premature. We therefore recommend expanding future evaluations to include more examples of this rare but chemically relevant class, enabling more robust statistical analysis of model performance on complex splitting patterns.

Finally, we note that the relatively lower precision observed for the `s` (singlet) class likely arises from both structural and statistical factors. From a structural standpoint, singlets represent the fundamental building blocks of all multiplet types, and ambiguous or simplified peaks in other patterns may resemble isolated singlets. As such, misclassifications involving the singlet class are expected to occur more frequently than for more distinct phenotypes. On the statistical side, singlets are sparsely represented in the annotated experimental set, with each instance contributing only a single labelled point. In contrast, multiplets from all other classes contribute multiple labelled points, since all spectral points lying between the outermost peaks of a multiplet are included in the annotation. This leads to an imbalance where singlets false positives accumulate more

rapidly than true positives, artificially lowering the computed precision.

Ultimately, these results establish that MuSe Net generalises effectively from synthetic to real-world NMR spectra, delivering high recall across most multiplet categories. By retaining robust performance in the presence of covariate shift and structural ambiguity, it proves to be a highly viable tool for the automated analysis of complex ^{1}H NMR spectra.

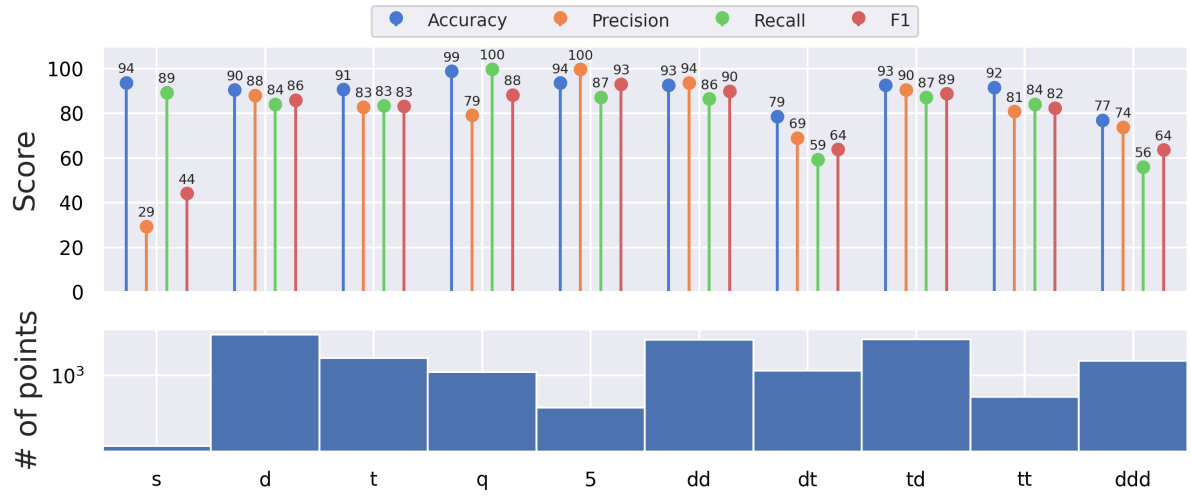


Figure 5.3: Evaluation metrics for point-by-point MuSe Net classification on the experimental test set (top panel), together with the ground truth distribution of spectral points across multiplet classes (bottom panel). he classes include: *s* for singlets, *d* doublets, *t* triplets, *q* quartets, 5 quintets, *dd* doublets of doublets, *dt* doublets of triplets, *td* triplets of doublets, *tt* triplets of triplets, and *ddd* doublets of doublets of doublets. The *dddd* class is not reported because only one such multiplet was present in the experimental test set. Scores are reported in percentages. The number of spectral points per class reflects the signal's span over the spectral range. Signal regions are annotated between the outer peaks of the multiplet: for singlets, only a single spectral point is labelled.

5.2.3 Prediction uncertainty

Incorporating uncertainty quantification into our deep learning framework for multiplet segmentation, MuSe Net, provides a principled way to assess the trustworthiness of model predictions and gain insight into what the model has actually learned. This approach not only enables users to interpret classification outputs more critically but also offers a mechanism to identify and analyse failure modes in the prediction process. Specifically, we employed two complementary uncertainty measures: the mutual information (MI), which reflects epistemic uncertainty (i.e., uncertainty due to limited knowledge or insufficient training data), and the *expected entropy* of the individual model predictions (H_2), which captures aleatoric uncertainty (i.e., irreducible noise or class ambiguity inherent to the data). The joint use of these two measures allows for a more nuanced understanding of prediction reliability, as their relative magnitudes and distributions can highlight whether misclassifications stem from data ambiguity or from gaps in the model's learned representation.

We began our analysis on the synthetic validation set, which was constructed using the same automated data generation pipeline as the training set. Because these synthetic

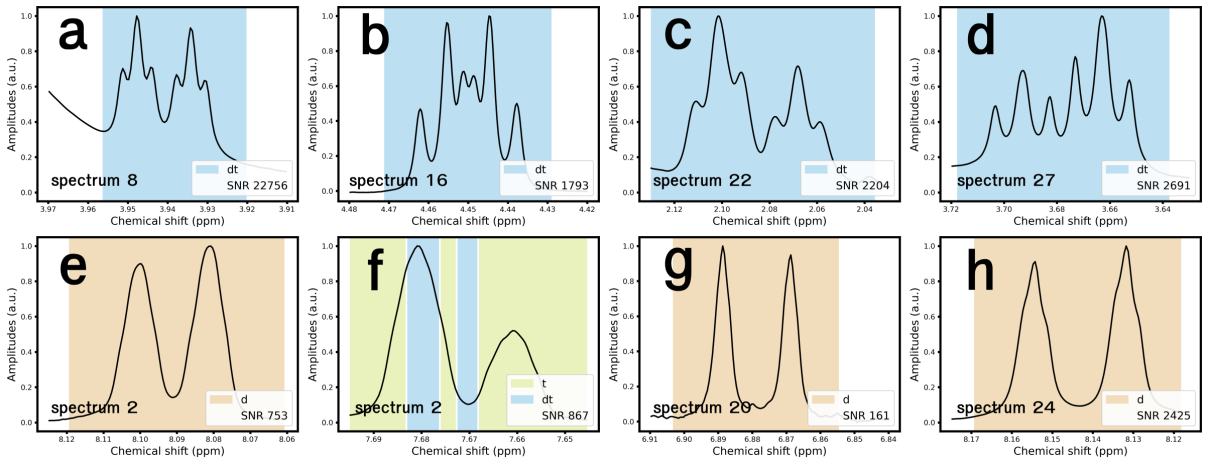


Figure 5.4: MuSe Net predictions on selected doublets of triplets from the experimental test set. **Top row**: the algorithm successfully classified challenging cases such as doublets of triplets appearing on the shoulder of larger peaks (**a**), exhibiting non-standard phenotypes (**b**), or affected by distortions (**c**, **d**). **Bottom row**: errors typically occurred in instances where the multiplet peaks overlapped significantly, making clear peak separation and identification difficult.

samples are drawn from a distribution that matches the training distribution, we expect the dominant source of uncertainty to be aleatoric rather than epistemic. That is, prediction errors in this context are most likely due to overlapping class features or visually similar multiplet structures being mapped to the same region in the feature space, rather than to model ignorance. To test this hypothesis, we compared the distributions of MI and H_2 across correctly and incorrectly classified samples using boxplots (see Figure 5.6 and Figure 5.7). As expected, both metrics exhibit higher values for incorrect classifications, reflecting increased model uncertainty when the prediction is wrong. However, we observed that H_2 increases significantly more than MI in these cases, suggesting that incorrect predictions in the synthetic domain are primarily driven by high aleatoric uncertainty. This makes expected entropy a useful diagnostic tool: high values of H_2 serve as a strong indicator of likely misclassification. Since the model was trained and validated on data generated from the same distribution, epistemic uncertainty remains low, reinforcing the conclusion that prediction errors are not due to model unfamiliarity but rather to intrinsic ambiguities between multiplet classes.

An exception to this trend was observed in the case of `dddd` (doublet of doublet of doublets of doublets) patterns. For this class, incorrect predictions showed a simultaneous rise in both H_2 and MI, implying that misclassification arises not only from class overlap but also from insufficient representation or understanding of this complex class in the training data. This observation points to a higher epistemic component in the model's uncertainty for `dddd`, potentially due to the limited number of such patterns in the training set or their higher structural variability.

We repeated this uncertainty analysis on the experimental test set to examine model behaviour under real-world conditions (see Figure 5.6 and Figure 5.8). In contrast to the synthetic data, both epistemic (MI) and aleatoric (H_2) uncertainty levels were elevated across the board, even for correctly classified multiplets. This suggests that real-world spectral data introduces a broader range of variability, such as phase distortions, unexpected baseline shifts, or unmodeled interactions, that the model has not encountered

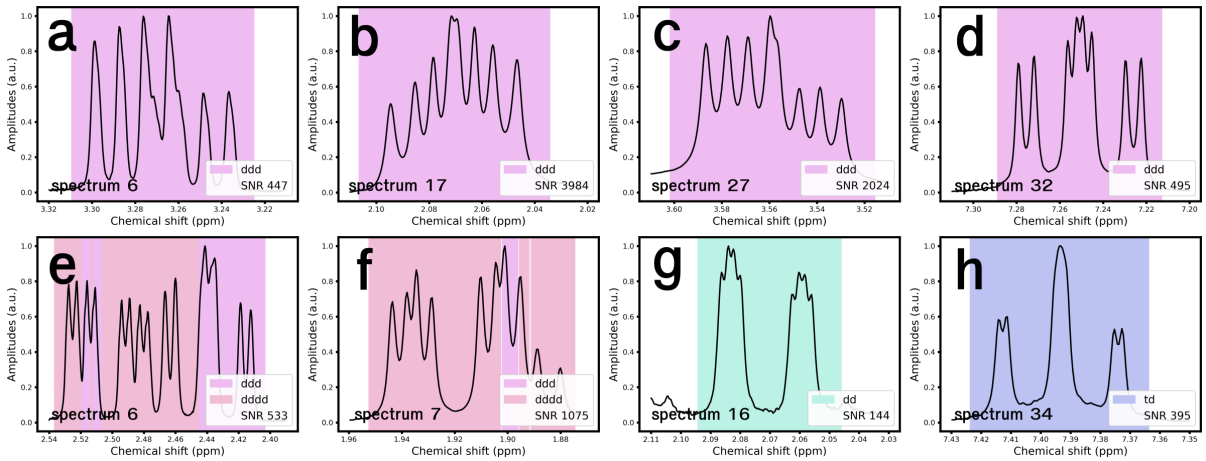


Figure 5.5: MuSe Net predictions on a selection of doublets of doublets of doublets (ddd) from the experimental test set. **Top row**: the algorithm successfully recognized ddd patterns across a range of cases featuring different combinations of coupling constants and line widths, resulting in varied peak configurations that all correspond to the same underlying splitting. **Bottom row**: misclassifications included two adjacent ddd mistaken for dddd (**e**), a ddd located near two additional peaks also misidentified as dddd (**f**), a ddd with one dominant coupling constant interpreted as two dd (**g**), and a ddd with four central merging peaks flanked by side peaks incorrectly classified as a td (**h**).

during training. Most notably, on misclassified experimental samples, MI increased more substantially than H_2, indicating that discrepancies between the experimental and synthetic domains are now a major source of prediction failure. This shift in the dominant uncertainty source from aleatoric to epistemic highlights a distributional mismatch and reflects the limitations of training exclusively on synthetic data.

In such cases, mutual information becomes especially valuable, as it can guide the NMR practitioner in identifying spectral regions or multiplet types where the model expresses genuine uncertainty about its own knowledge. By flagging predictions with high MI, the model can point the user toward instances that may benefit from additional verification or targeted refinement, such as reprocessing the spectrum, checking annotation consistency, or augmenting the training set with similar real-world examples. Overall, this uncertainty-aware framework not only enhances the transparency and interpretability of MuSe Net's predictions but also supports more informed and resilient decision-making in practical NMR analysis workflows.

5.2.4 Out Of Distribution detection

Out-of-distribution (OOD) detection in the context of multiplet segmentation can be naturally formulated as a binary classification task layered atop the primary segmentation objective. The goal is to discriminate between in-distribution (ID) and OOD inputs by leveraging the uncertainty estimates produced by the model—specifically, by identifying instances for which the predictions may be unreliable due to insufficient familiarity with the input features. In our implementation, this distinction is operationalised through a decision threshold θ_{OOD} applied to the mutual information $MI(x)$. Predictions associated with low uncertainty, that is when $MI(x) < \theta_{OOD}$, are retained as reliable ensemble outputs. Conversely, predictions for which $MI(x) > \theta_{OOD}$ are flagged as uncertain, and

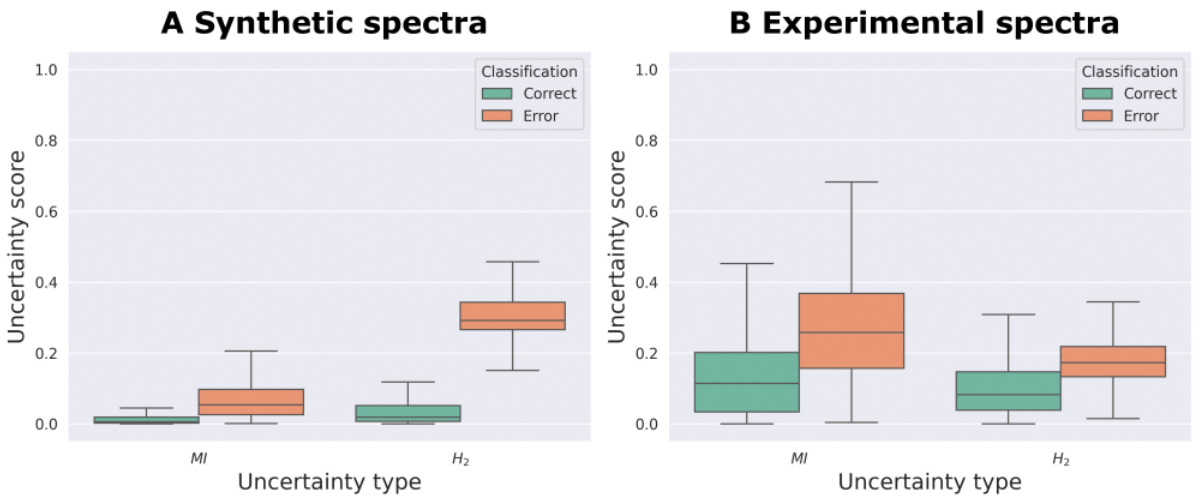


Figure 5.6: Mutual information and expected entropy of the individual models within the ensemble for correct and incorrect classifications on the synthetic validation set (A) and the experimental test set (B).

the corresponding inputs are conservatively labelled as generic, non-specific multiplets. This thresholding mechanism provides a straightforward yet effective means of filtering potentially unreliable predictions and enhancing model robustness under domain shift.

The quality of this OOD detection strategy fundamentally depends on the separability between ID and OOD samples in the model's learned feature space, and on the extent to which this separation is reflected in the model's uncertainty estimates. In ideal conditions, where the features of OOD inputs diverge meaningfully from those of the training distribution, mutual information provides a reliable proxy for this divergence. To empirically assess this, we analysed the distribution of MI values for ID and OOD multiplets, visualising their spread and degree of overlap (see Figure 5.9). As anticipated, MuSe Net produces noticeably higher uncertainty scores in response to OOD inputs, confirming the model's ability to recognise when it encounters unfamiliar or atypical spectral patterns.

Ideally, the uncertainty distributions associated with ID and OOD inputs, denoted by $P_{ID}(MI)$ and $P_{OOD}(MI)$, respectively,would be entirely disjoint, enabling perfect discrimination via an optimally chosen threshold θ_{OOD}. However, in real-world applications, the two distributions often exhibit partial overlap, meaning that intermediate values of mutual information may be observed for both ID and OOD samples. The degree of this overlap directly impacts the difficulty of setting a reliable threshold: the more the distributions coincide, the greater the risk of false positives (misclassifying ID samples as OOD) or false negatives (failing to detect true OOD samples). In contrast, a larger separation between the two distributions makes threshold selection more robust and improves the reliability of OOD detection.

To quantify the similarity between $P_{ID}(MI)$ and $P_{OOD}(MI)$, we employed the Weitzman overlapping coefficient, a statistical metric introduced in [113] that measures the extent of distributional overlap. This scalar quantity offers a concise and interpretable summary of the detection landscape: a lower overlap coefficient corresponds to greater separability between ID and OOD uncertainty profiles and, therefore, to more reliable uncertainty-based filtering. Overall, our results demonstrate that mutual information serves as an effective indicator of distributional shift in the multiplet classification setting and highlight MuSe Net's capacity to self-diagnose when confronted with unfamiliar

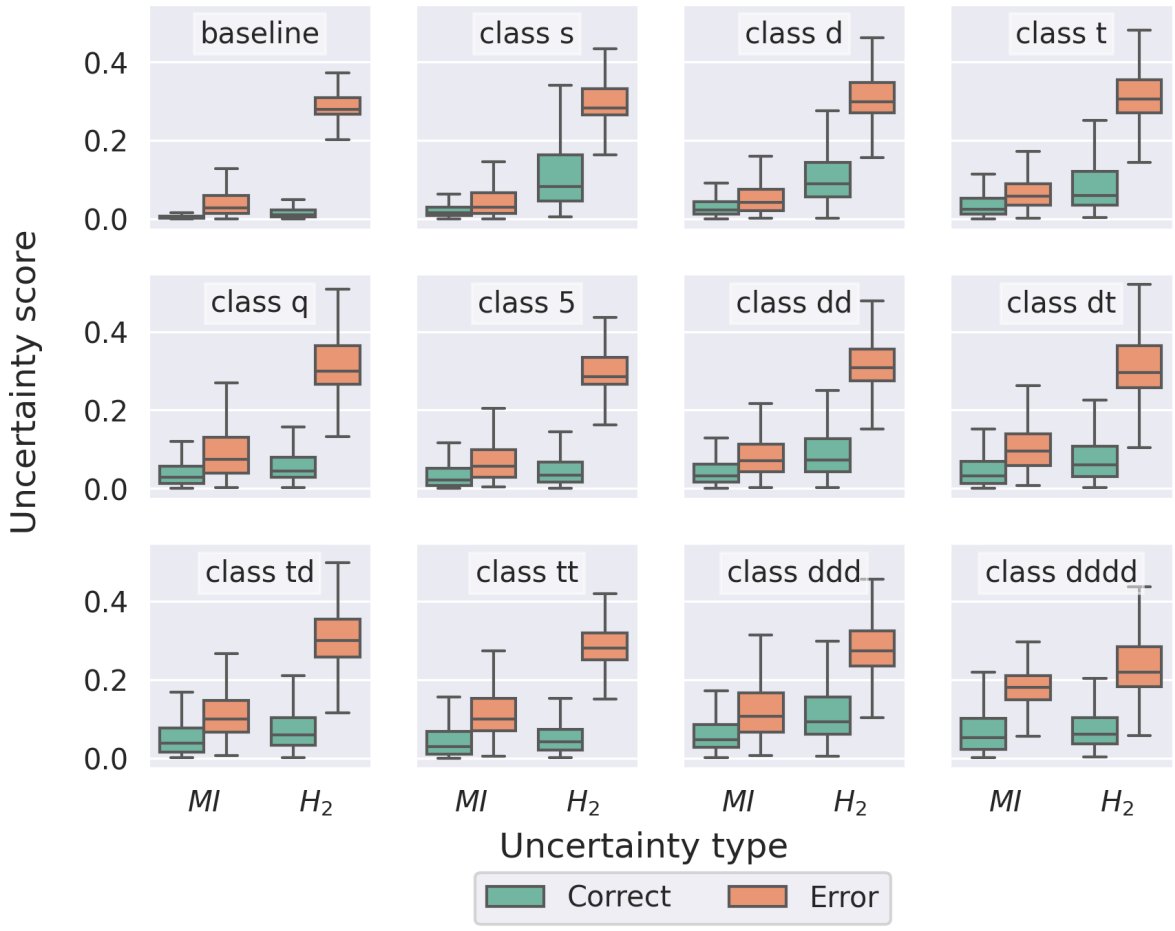


Figure 5.7: Mutual information and expected entropy of individual ensemble members for correctly and incorrectly classified samples, analyzed per class, on the synthetic validation set.

signal configurations.

5.2.5 Performance as a function of signal to noise ratio

A variety of factors can compromise the signal-to-noise ratio (SNR) in 1D ^{1}H NMR spectroscopy. Signal quality is often degraded by inadequate metabolite concentrations, poor solvent suppression, spectral distortions, and the presence of impurities. Additionally, line broadening, typically driven by chemical exchange events, sample inhomogeneity, or magnetic field irregularities, can severely diminish the SNR. At the same time, while ultra-high field systems remain the gold standard, accessible and compact benchtop spectrometers are gaining widespread adoption. This convenience, however, inherently limits sensitivity. Consequently, it is vital that automated NMR analytical algorithms perform reliably even when processing low-SNR spectra.

To assess the resilience of the MuSe Net framework under these suboptimal conditions, we systematically injected progressive amounts of white noise into our experimental test data and re-evaluated the model. As illustrated in Figure 5.10, which maps classification accuracy against the average multiplet SNR (omitting water signals), the predictive performance remained remarkably stable even at severely degraded noise levels. Notably, in

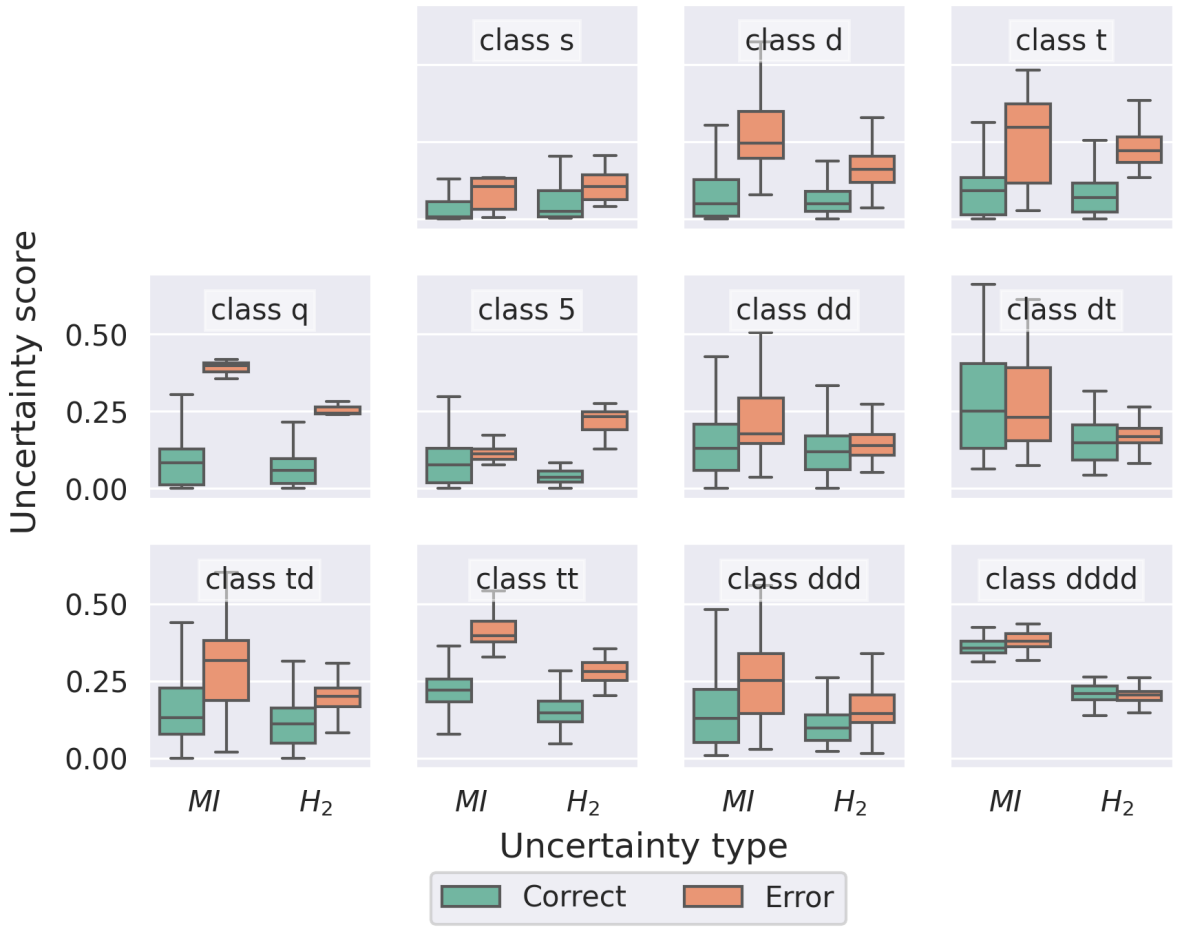


Figure 5.8: Mutual information and expected entropy of the ensemble's individual models for correctly and incorrectly classified instances, evaluated per class on the experimental test set.

several instances, introducing uncorrelated noise actually yielded a significant improvement in classification accuracy. We hypothesize that this seemingly paradoxical outcome occurs because minor additions of white noise obscure deceptive artifacts and spurious correlations generated during data acquisition and processing, thereby streamlining the algorithm's pattern recognition process.

To evaluate uncertainty quantification, Figures 5.11 and 5.12 show the distributions of the expected entropy (H_2) and mutual information (MI) for both in-distribution (ID) and out-of-distribution (OOD) multiplets relative to the average SNR. Because the median values remained largely stable as the noise escalated, we can conclude that MuSe Net provides highly consistent uncertainty metrics regardless of signal quality. We did observe, however, an overall decrease in uncertainty metrics (both H_2 and MI) for certain spectra where the average $SNR < 10$. This drop in uncertainty is entirely logical: as overwhelming noise masks substantial signal areas, these regions effectively mimic baseline characteristics. Since the algorithm is highly proficient at identifying these multiplet-free baseline regions, it confidently evaluates these noise-saturated sections with low uncertainty estimates.

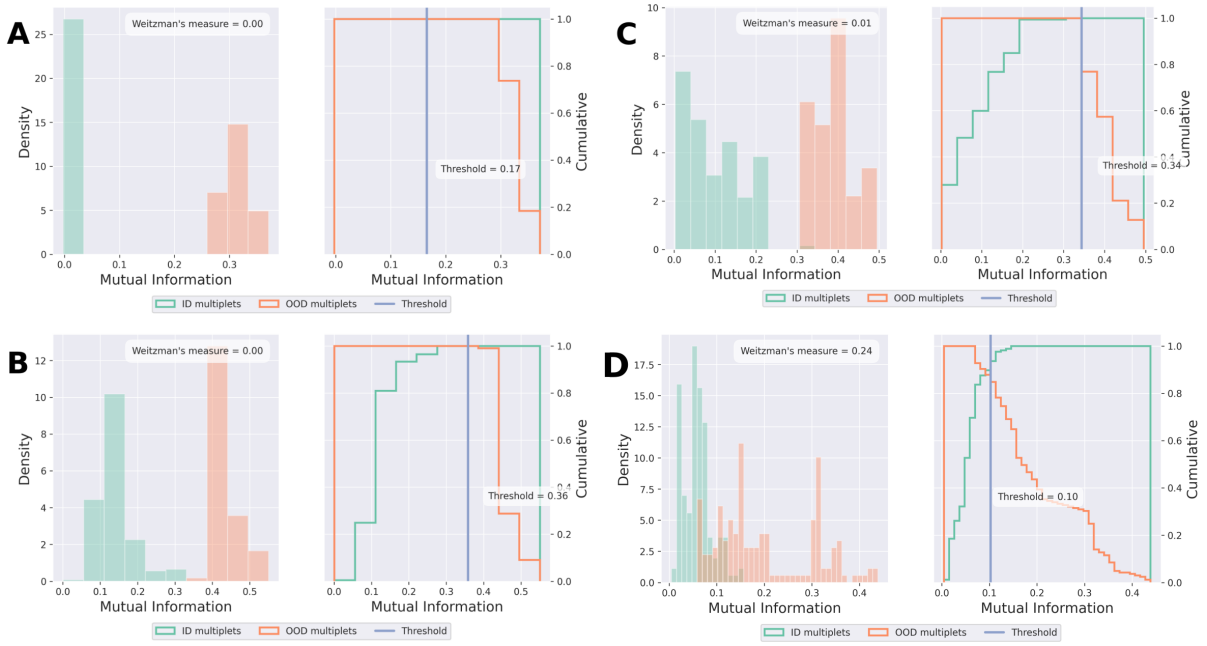


Figure 5.9: Out-of-distribution detection for the spectrum of Diethyl carbonate in DMSO (**A**), 2-phenylethanol in DMSO (**B**), Diaceton-D-galactopyranose in $CDCl_3$ (**C**), and 2,3:5,6-Di-o-isopropylidene-α-D-mannofuranose in DMSO (**D**). Left panel: mutual information distributions for in-distribution (ID) and out-of-distribution (OOD) samples using Weitzman's measure. Right panel: empirical cumulative distribution function (ECDF) of mutual information for ID samples, and one minus the ECDF for OOD samples, indicating the optimal threshold.

5.2.6 MuSe Net predictions

To illustrate the classification performance achieved by our method, we present a qualitative analysis of ^{1}H NMR spectra from compounds that vary in both structural complexity and experimental conditions. This selection showcases MuSe Net's flexibility across a range of real-world use cases.

We begin with the ^{1}H NMR spectrum of diethyl carbonate (see Figure 5.13). The sample was diluted in dimethyl sulfoxide (DMSO), and the spectrum was acquired at 400 MHz and processed to a spectral resolution of 0.22 Hz, yielding a high signal-to-noise ratio (SNR) of approximately 6×10^4. MuSe Net successfully detected and classified all the principal proton resonances. Notably, carbon satellites—minor signals caused by scalar coupling between ^{1}H nuclei and neighbouring ^{13}C nuclei—as well as the well-known five-peak DMSO resonance at 2.50 ppm, were both identified correctly and assigned to the general multiplet class m. Crucially, the DMSO resonance was not misclassified as a quintet, despite its five-peak appearance. This is because its intensity ratio of $1:2:3:2:1$ differs from the expected Pascal's triangle distribution $1:4:6:4:1$ of a true quintet, which arises in the case of four equivalent couplings. This demonstrates that MuSe Net effectively internalises multiplet phenotypes by simultaneously considering peak count and relative intensities, even for resonances resulting from a single coupling interaction, rather than relying on simplified pattern-matching heuristics.

MuSe Net also performs reliably in spectra of more structurally complex compounds with numerous chemically distinct hydrogen environments. For example, in Figure 5.14,

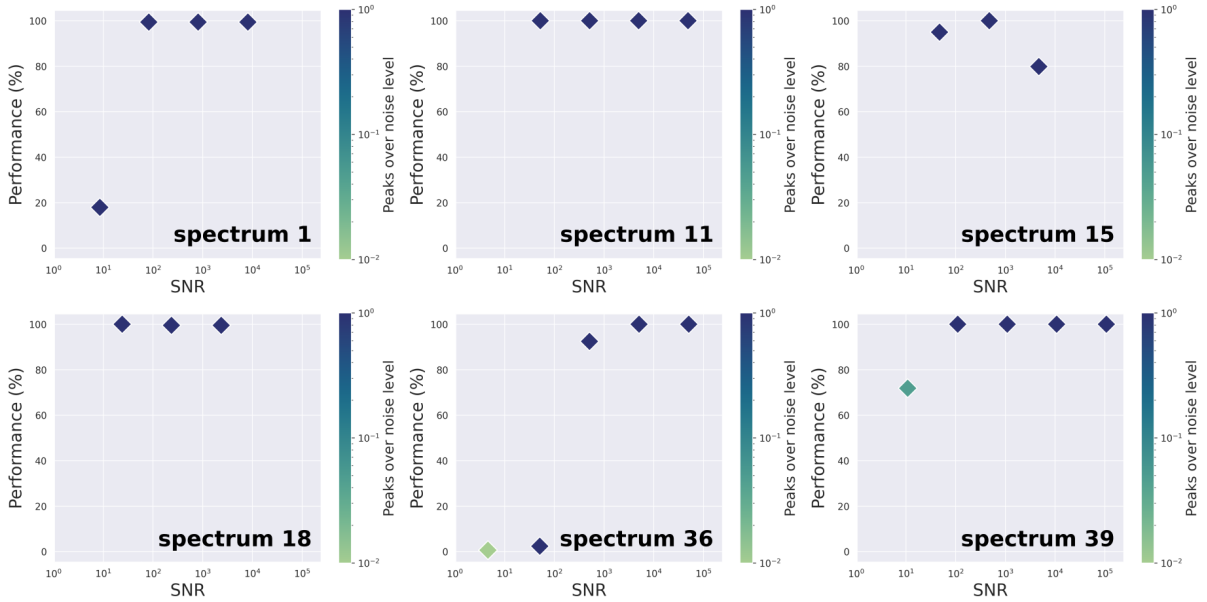


Figure 5.10: Classification accuracy as a function of signal-to-noise ratio for 6 ^{1}H NMR spectra from the experimental test set. Point color indicates the proportion of peaks above the noise threshold.

we show the model's predictions for the ^{1}H NMR spectrum of 2-phenylethanol dissolved in DMSO. This spectrum was acquired at 400 MHz and processed to a resolution of 0.24 Hz, with an SNR of approximately 6×10^3. MuSe Net correctly identified and classified all signal regions, including more challenging splitting patterns such as the triplet of doublets around 3.6 ppm. Furthermore, the model showed strong precision in resolving multiplet boundaries, accurately distinguishing noise from signal, and correctly segmenting overlapping but distinct neighbouring resonances. This ability is exemplified in inset box **d** of Figure 5.14, where two closely spaced multiplets are delineated with high fidelity.

Finally, we evaluated MuSe Net on the spectrum of diacetone-D-galactose dissolved in deuterated chloroform ($CDCl_3$), displayed in Figure 5.15. The spectrum was recorded at 400 MHz and processed to a resolution of 0.22 Hz, with an SNR of roughly 7×10^3. Once again, the model achieved highly accurate segmentation and classification. In particular, the residual solvent peak for $CDCl_3$ at 7.26 ppm (see inset **a** of Figure 5.15) was predominantly classified as a singlet but also partially assigned to the general multiplet class. This hybrid classification likely reflects increased uncertainty in the model's prediction, consistent with the slightly lower SNR and the fact that such residual signals often exhibit irregularities due to contamination or instrumental artefacts. Indeed, MuSe Net appears to capture this variability: signal regions with ambiguous morphology or low SNR, typical of impurities, residues, or acquisition artefacts, are frequently tagged as general multiplets with correspondingly elevated uncertainty scores (see inset **d** of Figure 5.15).

Altogether, these examples demonstrate MuSe Net's robustness across a spectrum of use cases, including high-SNR spectra with sharp peaks, chemically complex molecules with multiple coupling patterns, and lower-SNR regions prone to artefacts. The model not only classifies multiplets accurately but also expresses calibrated confidence that reflects the underlying quality and ambiguity of each signal region.

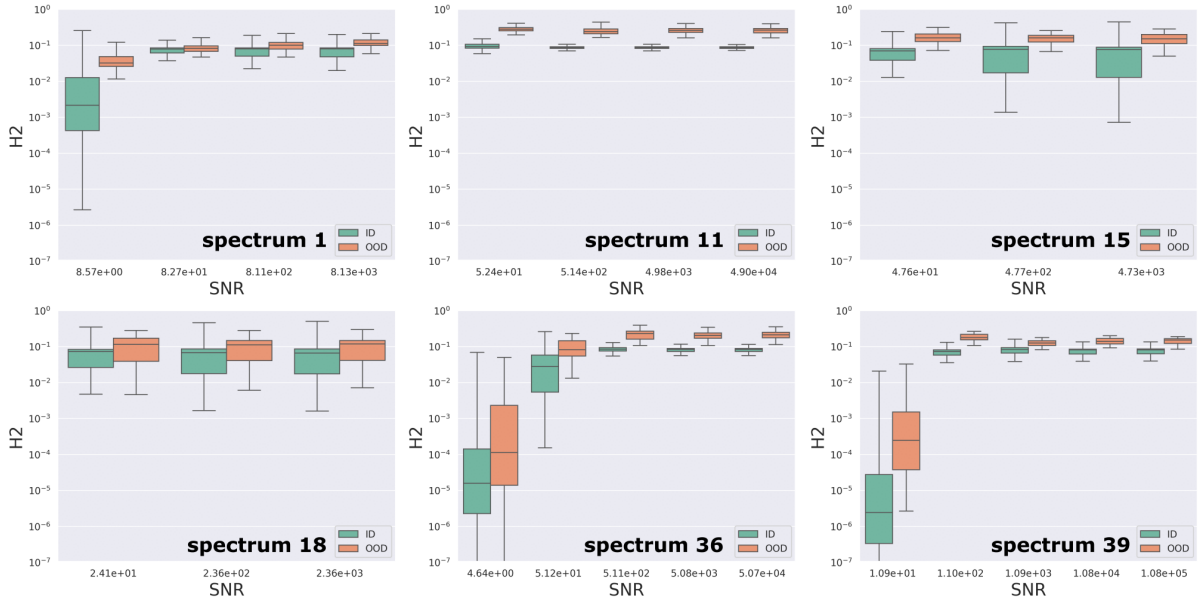


Figure 5.11: Expected entropy distribution of individual models on ID and OOD multiplets as a function of signal-to-noise ratio for 6 ^{1}H NMR spectra from the experimental test set.

5.3 From multiplet classification to structure verification

As an illustrative example of the use of MuSe Net predictions beyond the task of multiplet classification, we describe a procedure for structure verification applied to the spectrum of diethyl carbonate shown in Fig. 5.13. In this setting MuSe Net provides prior information on the spin system, which is subsequently employed within a Bayesian spectral fitting framework [114] developed in the context of the Innosuisse collaboration with Bruker Switzerland AG.

The Bayesian algorithm requires as input the number and type of spins present in the molecule together with estimates of their chemical shifts and spin–spin coupling constants. Optionally, an estimate of the linewidth can also be supplied in order to further constrain the fitting procedure. To connect the MuSe Net output with the Bayesian model we implemented a simple pipeline that converts the predicted segmentation of the spectrum into suitable prior distributions.

First, from the MuSe Net segmentation we determine the spectral ranges corresponding to each multiplet by grouping consecutive spectral points assigned the same class label. The chemical shift of each multiplet is then estimated as the midpoint of the corresponding spectral interval. Signals not associated with the main molecule are subsequently removed through a set of filtering rules. Residual solvent resonances are identified by comparing their chemical shifts with tabulated reference values and are therefore excluded from the analysis. Carbon satellite signals are removed by applying combined constraints on both their relative amplitudes and their characteristic positions with respect to the parent peak.

In the present example the underlying spin system is particularly simple. Due to molecular symmetry the proton spin system of diethyl carbonate reduces to an A_2X_3 configuration. The relative spin abundances are estimated by integrating the spectral

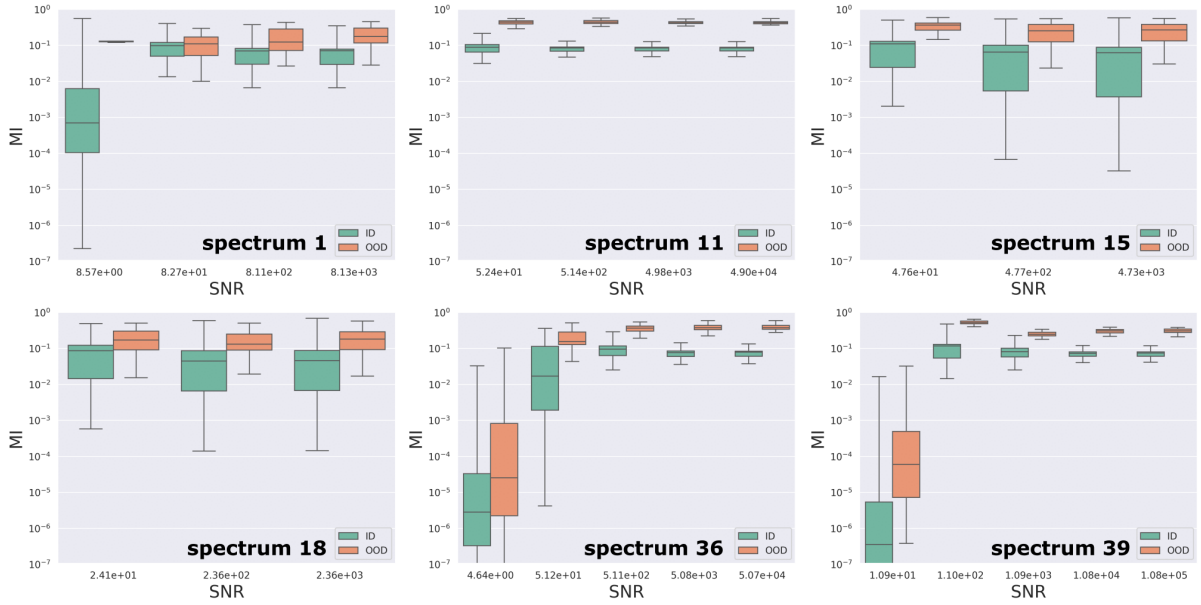


Figure 5.12: Mutual information distribution on ID and OOD multiplets plotted against signal-to-noise ratio for 6 ^{1}H NMR spectra from the experimental test set.

intensity within each multiplet range obtained from the MuSe Net segmentation. The spin–spin coupling constants are determined from the positions of the individual peaks within each multiplet, which are extracted using the `mldcon` peak detection algorithm and interpreted according to the predicted splitting pattern. MuSe Net does not provide explicit information on which resonances are mutually coupled. However, it is often possible to infer plausible coupling relationships by identifying identical coupling constants appearing in different multiplets. This strategy allows a consistent assignment of couplings in simple spin systems such as the present case. Nevertheless, this approach relies on heuristic inference and remains feasible only when the spectrum is sufficiently sparse. More systematic strategies would be required in the presence of strongly overlapping multiplets or more complex spin topologies. Finally, an estimate of the linewidth is also obtained from the `mldcon` procedure.

As shown in Table 5.2 and in the reconstructed spectrum (Fig. 5.16), the priors derived from the MuSe Net predictions provide sufficiently accurate initial conditions for the Bayesian fitting algorithm. The resulting spectral reconstruction is consistent with the expected structure of the molecule, demonstrating that MuSe Net outputs can be effectively integrated into automated workflows for structure verification.

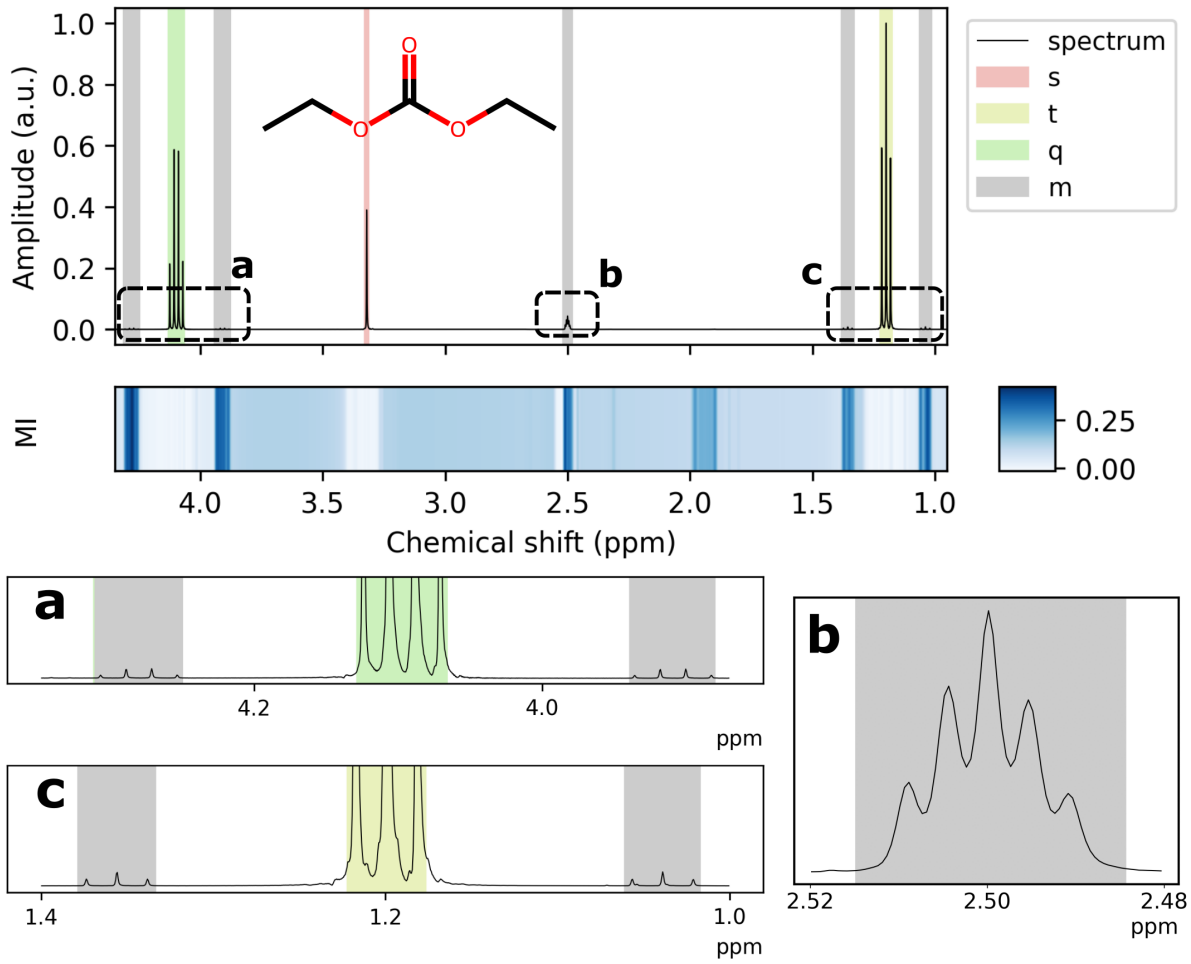


Figure 5.13: MuSe Net classification performance on the ^{1}H NMR spectrum of diethyl carbonate in DMSO: top panel shows the spectrum (black curve) with predicted multiplet classes (shaded regions), where s indicates singlets, t triplets, q quartets, and m general multiplets; middle panel displays mutual information as a continuous colour gradient, with darker tones representing higher uncertainty. Inset boxes a, b, and c show, respectively, the quartet at 4.1 ppm with its carbon satellites, the five-peak DMSO resonance at 2.5 ppm, and the triplet at 1.2 ppm with its carbon satellites.

	Chemical shift (ppm / Hz)	Width (Hz)	J-coupling (Hz)
Prior	CH_2 : 4.10/1639 CH_3 : 1.20/480	CH_2 : 0.68 CH_3 : 0.73	CH_2 : 7.13 CH_3 : 7.15
Posterior	CH_2 : 4.10/1640 CH_3 : 1.20/480	CH_2 : 0.60 CH_3 : 0.68	CH_2 : 7.11 CH_3 : 7.11

Table 5.2: Prior and posterior estimates of the spectral parameters used in the Bayesian spectral fitting procedure. The prior values are extracted automatically from the MuSe Net multiplet segmentation together with peak positions and linewidth estimates obtained from the `mldcon` peak detection algorithm. The posterior values correspond to the parameter estimates obtained after the Bayesian fitting algorithm has been applied to the experimental spectrum.

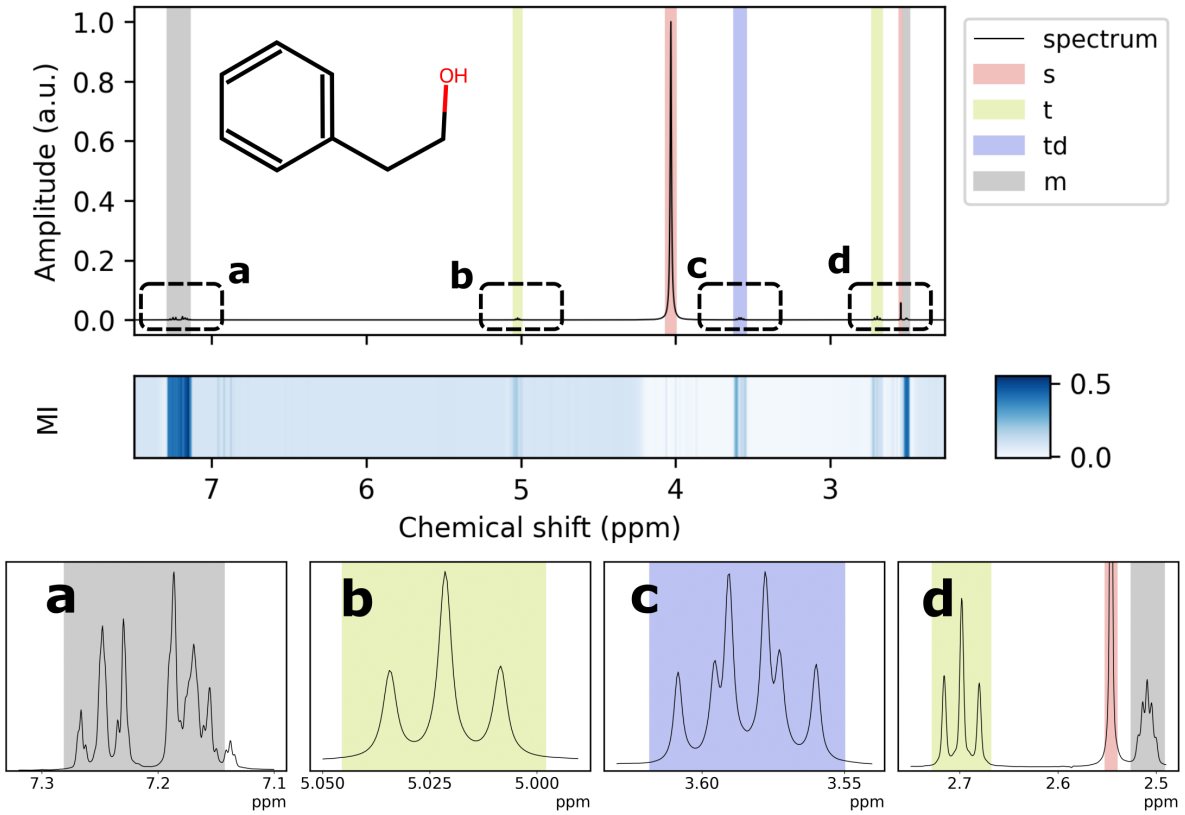


Figure 5.14: MuSe Net classification performance on the ^{1}H NMR spectrum of 2-phenylethanol in DMSO: top panel shows the spectrum (black curve) with predicted multiplet classes (shaded regions), where *s* indicates singlets, *t* triplets, *td* triplet of doublets, and *m* general multiplet; middle panel displays mutual information as a continuous colour gradient, with darker tones representing higher uncertainty. Inset boxes a, b, c, and d highlight a second-order quartet and triplet at 7.16 and 7.25 ppm, a triplet at 5.02 ppm, a triplet of doublets at 3.58 ppm, and a triplet, singlet, and five-peak DMSO resonance at 2.70, 2.55, and 2.50 ppm, respectively.

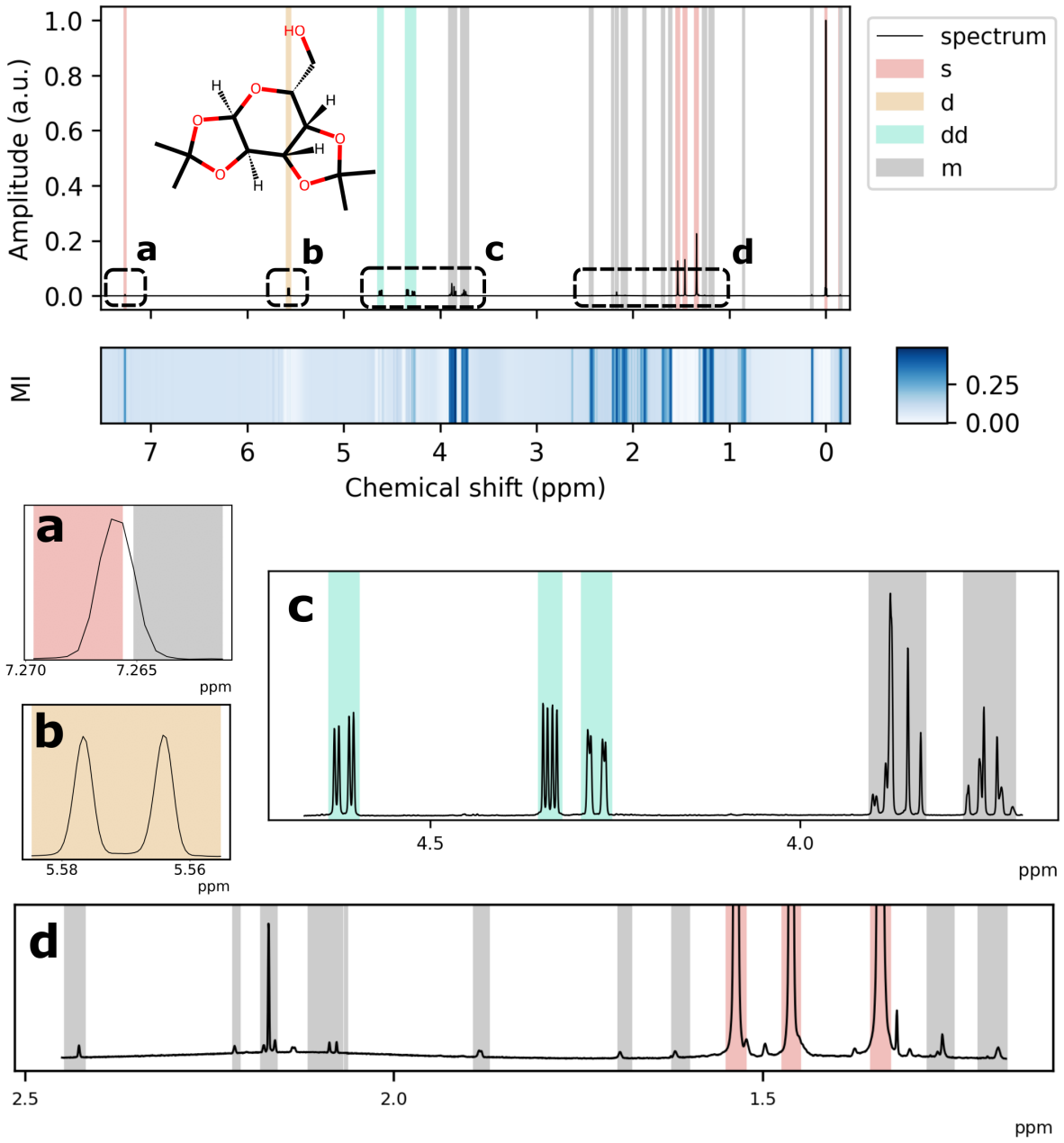


Figure 5.15: MuSe Net classification performance on the ^{1}H NMR spectrum of diacetone-D-galactose in $CDCl_3$. Top panel: the spectrum (black curve) with predicted multiplet classes (shaded regions), where *s* indicates singlets, *d* doublets, *dd* doublets of doublets, and *m* general multiplet. Middle panel: mutual information shown as a continuous colour gradient, with darker tones indicating higher uncertainty. Inset boxes a, b, c, and d show: (a) the residual $CDCl_3$ peak at 7.26 ppm, (b) a doublet at 6 ppm, (c) three doublets of doublets and two overlapping multiplets around 4 ppm, and (d) three singlets with small artefact peaks around 2 ppm.

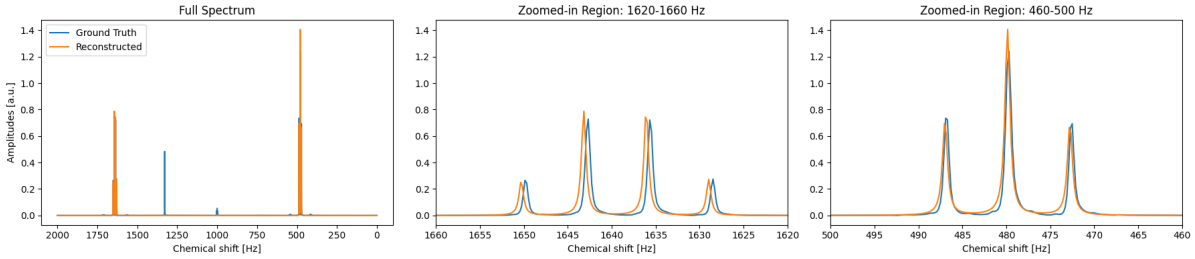


Figure 5.16: Comparison between the experimental spectrum (blue) and the reconstructed spectrum obtained through the Bayesian fitting algorithm (orange). The priors used for the reconstruction are automatically extracted from the MuSe Net multiplet segmentation together with peak positions and linewidth estimates provided by the `mldcon` peak detection algorithm.

Chapter 6

Improving sensitivity: photo-CIDNP NMR

6.1 Intrinsic Sensitivity Limitations of NMR

Nuclear magnetic resonance (NMR) spectroscopy derives its distinctive analytical power from the exceptionally long transverse relaxation times (T_2) over which nuclear spins can sustain coherent, observable magnetization following excitation. In liquid-state NMR, transverse spin coherence can persist for even a few seconds, enabling spectral resolution on the order of hertz or better. This high frequency resolution allows not only the extraction of detailed spectral fingerprints but also access to slow dynamical processes, including translational and rotational diffusion, conformational rearrangements, and chemical or physical transformations occurring on relatively long time scales.

These favorable relaxation properties originate primarily from the weak interaction between nuclear spins and their molecular environment. The magnetic dipole moments of nuclei with nonzero spin are intrinsically small, several orders of magnitude smaller than that of the electron, and consequently couple only weakly to fluctuating local magnetic fields arising from molecular motion. As a result, both longitudinal (T_1) and transverse (T_2) relaxation processes proceed slowly, thereby preserving signal coherence and enabling high-resolution spectroscopy.

However, the same weak spin–field interaction that underpins the strengths of NMR also constitutes its principal limitation. Because nuclear magnetic moments interact only weakly with both static and oscillating magnetic fields, the signals generated in NMR experiments are inherently weak. In contrast to spectroscopic techniques probing electronic, vibrational, or rotational transitions, such as ultraviolet–visible, infrared, Raman, or microwave spectroscopy, NMR involves transitions between nuclear spin states separated by extremely small energy gaps. The associated radiofrequency absorption or emission is therefore detected with orders of magnitude lower sensitivity than higher-energy spectroscopic modalities.

At the microscopic level, the origin of this limitation can be understood by considering an ensemble of nuclei with nonzero spin placed in a static magnetic field. The applied field induces a partial alignment of nuclear spins, giving rise to a net nuclear magnetization. Yet, because the Zeeman interaction energy is small compared to the thermal energy at ambient temperature, the magnetic field competes only weakly with thermal randomization. Consequently, the orientational preference of nuclear spins is minimal.

This effect is quantitatively described by the nuclear spin polarization [115]. For an ensemble of isolated spin-$\frac{1}{2}$ nuclei, the two eigenvalues of the $z-$component of the spin operator $\hat{I}_z$ are $m_z = \pm 1/2$. The ground state $m_z = 1/2$ will be denoted by $|\alpha>$ and the higher energy state $m_z = -1/2$ will be denoted by $|\beta>$, while the populations of the two states will be respectively $p_{|\alpha>}$ and $p_{|\beta>}$. The sum of the populations is normalized to unity $p_{|\alpha>} + p_{|\beta>} = 1$. The net polarization defined as the difference in population between the two Zeeman states:

$$P = p_{|\alpha>} - p_{|\beta>} = 1 - 2p_{|\beta>}. \tag{6.1}$$

Once thermal equilibrium has been reached the relative populations of the two energy levels follow the Boltzmann distribution:

$$\frac{p_{|\alpha>}}{p_{|\beta>}} = e^{-\hbar\gamma B_0/k_B T}, \tag{6.2}$$

where γ is the gyromagnetic ratio of the nucleus, B_0 is the strength of the applied static magnetic field, k_B is the Boltzmann constant, and T is the absolute temperature. With this, the net polarization becomes:

$$P = \tanh\left(\frac{\hbar\gamma B_0}{2k_B T}\right). \tag{6.3}$$

Under thermal equilibrium conditions, the polarization depends on the strength of the applied magnetic field, the temperature, and the gyromagnetic ratio of the nucleus. At room temperature in the magnetic fields of modern high-resolution NMR spectrometers, the thermal polarization of proton spins is typically on the order of 10^{-4} to 10^{-5}, and even lower for nuclei with smaller gyromagnetic ratios.

Since the observable NMR signal intensity is directly proportional to this net spin polarization, the small population imbalance imposed by thermal equilibrium constitutes a fundamental sensitivity bottleneck.

Within the constraints of thermal equilibrium, numerous strategies have been developed to improve the signal-to-noise ratio (SNR) in magnetic resonance experiments. A major conceptual advance was the introduction of pulsed Fourier-transform NMR, which exploits the multiplex advantage to enable more efficient signal averaging. Further improvements have been achieved through technological developments in radiofrequency detection, including the design of more efficient inductive probes, detector miniaturization, and systematic noise reduction by means of cryogenic cooling of probes and signal preamplifiers. These approaches act primarily on the detection side of the experiment, enhancing sensitivity by reducing electronic noise rather than by increasing the intrinsic nuclear magnetization.

In parallel, general strategies aimed at increasing the equilibrium spin polarization itself have also been pursued. Because thermal polarization scales inversely with temperature and linearly with magnetic field strength, both sample cooling and stronger static magnetic fields increase the net equilibrium magnetization. The benefit to detection is even more pronounced, as the signal-to-noise ratio scales superlinearly with field strength (approximately as $B_0^{3/2}$ to B_0^2), owing to the additional dependence of the induced signal on the nuclear precession frequency. In practice, however, these approaches are subject to significant limitations. Substantial sample cooling is not always compatible with the

system under investigation or the desired experimental conditions, while the development of NMR magnets with ever-increasing field strength follows a slow, technologically demanding, and economically costly trajectory.

As a result, although these technological and methodological advances have played a crucial role in extending the applicability of NMR spectroscopy, further progress along these lines is expected to be incremental. Achieving substantial additional sensitivity gains therefore calls for fundamentally different approaches that circumvent the limitations imposed by thermal equilibrium polarization.

6.2 Defining Hyperpolarization

In Chapter 2, it was noted that spin polarization, defined by an imbalance in the populations of nuclear spin energy levels, produces a net longitudinal magnetization in an external magnetic field, which constitutes the source of the detectable NMR signal upon radiofrequency excitation [75, 72].

From a mathematical perspective, a spin system is said to be polarised when the density operator ρ describing its quantum state is anisotropic, meaning that it is not invariant under rotations of the reference coordinate frame. The magnitude of the polarization is therefore determined by the extent of this anisotropy. For an ensemble of isolated spin-$\frac{1}{2}$ nuclei, the density operator can be conveniently expanded as follows:

$$\rho = \sum_{\lambda=0}^{1} \sum_{\mu=-\lambda}^{\lambda} \rho_{\lambda\mu} \mathcal{T}_{\lambda\mu}, \tag{6.4}$$

where $\mathcal{T}_{\lambda\mu}$ denote irreducible spherical tensor operators, and $\rho_{\lambda\mu}$ are the corresponding expansion coefficients. The rank-0 coefficient represents the isotropic part of the density operator, while the rank-1 coefficients are directly related to the longitudinal polarization and therefore to the net magnetization along the direction of the external magnetic field.

Polarization can be quantified through a spin order parameter defined in terms of the von Neumann entropy of the spin system. Such parameter measures the deviation of the system from a maximally disordered state: it vanishes when the entropy is maximal and approaches unity as the system tends toward a pure quantum state:

$$\mathrm{SO} = \frac{S_{\mathrm{vN}}^{\max} - S_{\mathrm{vN}}}{S_{\mathrm{vN}}^{\max}}, \tag{6.5}$$

where the von Neumann entropy S_{vN} is given by

$$S_{\mathrm{vN}} = -\mathrm{Tr}\{\rho \ln \rho\}, \tag{6.6}$$

with ρ denoting the density operator. The maximum entropy,

$$S_{\mathrm{vN}}^{\max} = \ln N_{\mathrm{H}}, \tag{6.7}$$

corresponds to a completely mixed state with equal populations of the N_{H} quantum states and no coherences.

The von Neumann entropy can be used to formulate a general criterion for hyperpolarization. For a spin system in thermal equilibrium with an environment at temperature T, the von Neumann entropy is given by

$$S_{\mathrm{vN}}^{\mathrm{eq}}(T) = -\mathrm{Tr}\{\rho_{\mathrm{eq}}(T) \ln \rho_{\mathrm{eq}}(T)\}, \tag{6.8}$$

where $\rho_{\mathrm{eq}}(T)$ denotes the thermal equilibrium density operator.

On this basis, a system is defined as hyperpolarised if its von Neumann entropy satisfies

$$S_{\mathrm{vN}} < S_{\mathrm{vN}}^{\mathrm{eq}}(T), \tag{6.9}$$

where T is the temperature of the surrounding environment.

This entropy-based criterion defines hyperpolarization independently of explicit population differences or the presence of a net magnetic moment along a specific spatial direction.

While the entropy-based criterion provides a general and formally rigorous definition of hyperpolarization, a more qualitative and operational definition can be given by considering the practical motivation underlying spin hyperpolarization techniques. These approaches are primarily driven by the substantial signal enhancements they can provide in nuclear magnetic resonance experiments, particularly in situations where sensitivity is a limiting factor.

Accordingly, for an ensemble of isolated spin-$\frac{1}{2}$ nuclei, a hyperpolarised nuclear spin system can be defined operationally as a system prepared out of thermal equilibrium such that the resulting NMR signal intensity exceeds that obtained under thermal conditions. This requirement can be expressed by the condition $|p_{\mathrm{hyp}}| > |p_{\mathrm{therm}}|$.

The magnitude of enhancement required for effective application depends on the specific experimental context. Nevertheless, a commonly adopted operational criterion is that the increase in polarization, $|p_{\mathrm{hyp}}| - p_{\mathrm{therm}}$, exceeds the thermal equilibrium polarization p_{therm}, with $p_{\mathrm{therm}} > 0$. Under this condition, hyperpolarization results in at least a twofold enhancement of the NMR signal intensity relative to thermal equilibrium.

6.3 Hyperpolarization techniques

Several hyperpolarization strategies have been developed, including dynamic nuclear polarization (DNP) [63], parahydrogen-induced polarization (PHIP) [64], signal amplification by reversible exchange (SABRE) [65], and photochemically induced dynamic nuclear polarization (photo-CIDNP) [66, 67]. Dynamic nuclear polarization transfers high electron spin polarization from unpaired electrons to the target nuclei via microwave irradiation. The primary advantages of this approach include exceptionally high signal enhancements, often exceeding 10,000-fold. It is also a very general technique applicable to a vast array of molecules without requiring specific functional groups. However, DNP requires highly specialized equipment operating at ultra-low temperatures.

Parahydrogen-induced polarization is able to provide very high signal enhancements while being operationally much simpler and cheaper than DNP, bypassing the need for extreme cryogenic temperatures. On the other hand, since, it relies on the incorporation of parahydrogen into a target molecule via a catalytic hydrogenation reaction, PHIP requires a permanent chemical modification of the target molecule.

Signal amplification by reversible exchange represents an of parahydrogen-based hyperpolarization. It utilizes an organometallic catalyst that transiently and reversibly binds both parahydrogen and the target molecule. Similar to PHIP, SABRE is fast, cost-effective, and operates directly in the liquid state. The main limitations revolve around its heavy dependence on the chemical structure of the target. Additionally, the presence of the required metal catalyst in the final solution can complicate biological applications.

Photochemically induced dynamic nuclear polarization relies on the generation of transient radical pairs upon laser illumination of a photosensitizer dye. While exhibiting lower signal enhancements compared to other hyperpolarization methods, photo-CIDNP NMR offers several advantageous features. Its operational complexity is comparatively low, as polarization is triggered simply by illuminating the sample within the cryoprobe. Moreover, the experiment can be performed in aqueous solution at room temperature, without chemical modification of the protein, and is compatible with low micromolar concentrations of both ligand and protein (down to 5 μM of ligand and 2 μM of target). The experiment requires only a few seconds, as sufficient signal-to-noise ratio can often be achieved within a single scan. Its primary limitations are its modest absolute enhancement factors compared to other methods and its strict specificity, as it generally requires specific aromatic side chains (such as tryptophan or tyrosine). Projections suggest that approximately 25–30% of biologically significant small molecules are amenable to this technique. Consequently, this inherent specificity enforces a structural bias when compiling fragment screening libraries. Despite this restriction, recent evaluations of optimal chemical spaces for fragment-based screening demonstrate that a high degree of functional group variation can effectively offset a narrow range of core scaffolds [11]. Importantly, it has been established that introducing such functional variety merely modulates the photo-CIDNP signal without extinguishing the underlying hyperpolarization activity. In the following sections we will illustrate the quantum mechanical basis of the photo-CIDNP effect.

6.4 Quantum Mechanical Description of Photo-CIDNP effect

Radical pairs are the fundamental intermediates responsible for the chemically induced dynamic nuclear polarization (CIDNP) effect observed in NMR spectroscopy. A *radical pair* consists of two radicals, molecular species each possessing an unpaired electron, making them inherently paramagnetic, that are generated simultaneously and remain in close spatial proximity in solution. The unpaired electrons of the two radicals are quantum-mechanically correlated, meaning their spins are coupled in either a singlet or triplet spin state. This correlation has profound consequences for how these species evolve and recombine during chemical reactions.

When NMR spectroscopy is applied to monitor free radical reactions in solution, the resulting spectra can exhibit *non-Boltzmann* nuclear spin populations, leading either to opposite-sign components within a scalar-coupled multiplet, where some lines appear in absorption and others in emission (*multiplet effect*), or to uniform enhancement or inversion of an entire resonance (*net effect*). These enhancements arise because the nuclear spin state of magnetic nuclei can modulate the electronic spin dynamics of the radical pair through hyperfine interactions. In essence, the spin state of a nucleus can influence whether the radical pair reacts via the singlet or triplet pathway, thereby altering chemical yields and spin polarization.

The first observations of anomalously strong emissive NMR signals were reported by Bargon et al. [116] during their investigation of the thermally induced decomposition of dibenzoyl peroxide (DBPO) in solution. Nearly simultaneously, Ward and Lawler independently documented the same striking phenomenon [117]. Initially termed chemically induced dynamic nuclear polarization (CIDNP), this effect was first interpreted through

theories invoking the Overhauser mechanism and dynamic nuclear polarization, which emphasized cross-relaxation between electron and nuclear spins.

However, these early models failed to account for the magnitude of signal enhancements, often orders of magnitude beyond thermal equilibrium, and the coexistence of absorption and emission phases within the same nuclear resonance multiplet. Only subsequent theoretical frameworks based on the Radical Pair Mechanism (RPM), developed by Closs [118] and Kaptein [119], successfully explained these observations.

In the radical pair model, correlated radical pairs generated during the reaction undergo singlet-triplet intersystem crossing driven by hyperfine interactions and external magnetic fields. This interplay between nuclear and electronic spins leads to non-Boltzmann nuclear spin populations that manifest as net emission or enhanced absorption in the recombination or escape products. This breakthrough not only rationalized the characteristic phase patterns (Absorption/Emission rules) but also established CIDNP as a premier tool for probing radical reaction mechanisms, spin dynamics, and radical lifetimes in real time, offering deep insights into transient reactive intermediates, electron transfer processes, and detailed kinetic information often inaccessible by conventional NMR or optical spectroscopy.

6.4.1 Radical Pair Mechanism

Radical species emerge as correlated pairs when generated from non-radical precursors, typically in the singlet spin state characteristic of thermal reactions, or in the triplet state prevalent during photochemical excitations, and they recombine in pairs to yield non-radical products in the singlet configuration. This phenomenon appears to contradict the fundamental rule of spin conservation in elementary reactions, which dictates that triplet reactants can only produce triplet products, while singlet reactants yield singlet products. However, the formation of products exhibiting a spin multiplicity different from that of the initial reactants clearly demonstrates that intersystem crossing occurred within the reaction pathway. The radical pair mechanism provides one of the simplest explanations for this process, involving just three key elements: the initial reactants, a single intermediate (the radical pair) that experiences intersystem crossing, and the resulting products.

The radical pair mechanism is enabled by two key interactions: Zeeman interactions, characterized by the g-factors of the radicals, and hyperfine couplings, characterized by coupling constants a. The hyperfine interactions couple nuclear spins to the radical electrons, making the intersystem crossing (isc) rate strongly dependent on the specific nuclear spin configurations present in radicals R_1 and R_2. Certain nuclear spin states accelerate isc significantly (fast configurations), while others retard it (slow configurations).

If we consider radical pairs generated via the triplet channel, those containing fast isc nuclear spins preferentially convert to the singlet state, and the corresponding nuclear spins will be overrepresented in singlet pairs and underrepresented in the triplet pairs. Conversely, slow isc spins remain predominantly represented in the triplet pairs. This process constitutes effectively a nuclear spin sorting between the singlet and triplet states.

These nuclear spin population imbalances are preserved when radical pairs recombine into observable products, manifesting as dramatic NMR signal enhancements. Specifically, for any NMR transition connecting two nuclear spin states, overpopulation of the lower-energy spin state produces enhanced absorption, while overpopulation of the higher-energy spin state generates emission (negative amplitude signal).

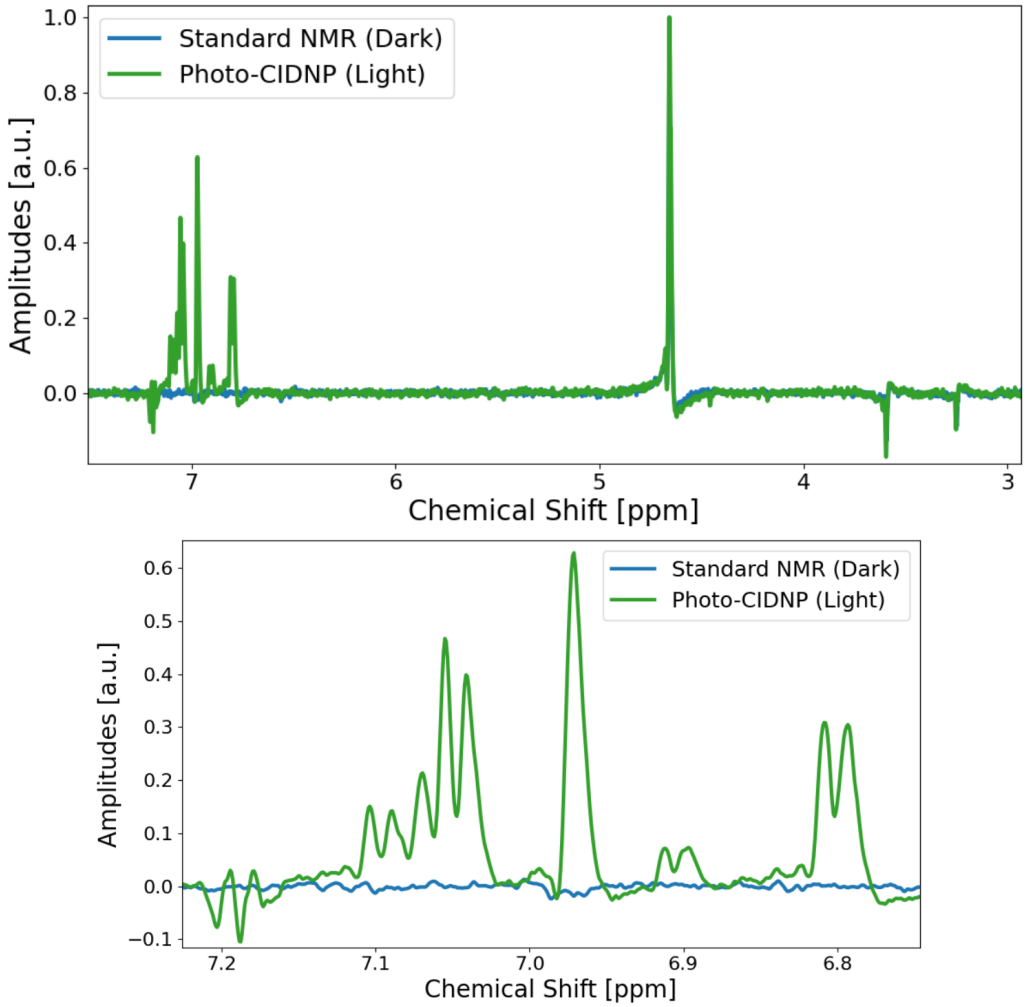


Figure 6.1: Comparison between the standard ^{1}H NMR spectrum and the corresponding photo-CIDNP NMR spectrum for one of the small molecules in the screening dataset (see Section 7.2.1). The upper panel shows the full spectral region, while the lower panel displays a magnified view (inset) of a selected region of the same spectrum to highlight the differences between dark and illuminated conditions.

Crucially, no actual nuclear spin flips occur during this mechanism, only selective sorting between states. This sorting is bidirectional: regardless of direction (diagonal routes: triplet → singlet or singlet → triplet), fast isc spins always enrich the product formed from the opposite state. Vertical routes (same-multiplicity: singlet → singlet or triplet → triplet) produce exactly opposite polarizations with respect to diagonal routes.

Overall, the radical pair mechanism responsible for CIDNP relies on three essential conditions: (i) the formation of radical pairs with a dominant electron spin multiplicity, (ii) nuclear-spin–dependent intersystem crossing between spin states, and (iii) unequal reaction rates for the competing recombination pathways.

6.4.2 The Spin Hamiltonian of a Radical Pair

The spin dynamics of a radical pair are described by the spin Hamiltonian $\hat{\mathcal{H}}$, which can be decomposed into three contributions,

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_z + \hat{\mathcal{H}}_{hf} + \hat{\mathcal{H}}_J, \tag{6.10}$$

where $\hat{\mathcal{H}}_z$ represents the Zeeman interaction of the unpaired electrons with the external magnetic field, $\hat{\mathcal{H}}_{hf}$ describes the hyperfine interactions between the electron and nuclear spins, and $\hat{\mathcal{H}}_J$ accounts for the exchange interaction between the two unpaired electrons. In solution, the radical pair undergoes rapid rotational diffusion and translational motion, such that the associated spin interactions can be treated as fully isotropic and therefore independent of spatial orientation or molecular alignment.

The magnitude of the Zeeman interaction between the electronic spins and the applied magnetic field B_0 is given by:

$$\hat{\mathcal{H}}_z = (g_1\hat{S}_{1z} + g_2\hat{S}_{2z})B_0\frac{\mu_B}{\hbar}, \tag{6.11}$$

where g_i is the isotropic g-factor of radical R_i, $\hat{S}_{iz}$ is the z-component of the electron spin operator on radical R_i, and μ_B is the Bohr magneton.

The CIDNP effect arises from hyperfine interactions between the spin angular momenta of the unpaired electrons ($\hat{S}$) and the nuclei ($\hat{I}$) within each radical:

$$\hat{\mathcal{H}}_{hf} = \sum_i a_i\hat{I}_i \cdot \hat{S}_1 + \sum_j a_j\hat{I}_j \cdot \hat{S}_2 \tag{6.12}$$

where a_i and a_j are the isotropic hyperfine coupling constants between nucleus i in radical R_1 and nucleus j in radical R_2 due to the Fermi contact interaction.

Depending on the overlap of the orbital wavefunctions, the two unpaired electrons in the spin-correlated radical pair may also experience an exchange interaction, described by the Hamiltonian term:

$$\hat{\mathcal{H}}_J = -J(r)\Big(\frac{1}{2} + \hat{S}_1 \cdot \hat{\mathcal{H}}_2\Big). \tag{6.13}$$

The distance dependent $J(r)$ function characterizes the strength of the interaction: it is generally assumed to decay exponentially and to be negligible beyond $1\,nm$.

When the radicals of a pair are in close contact, the exchange interaction dominates the spin Hamiltonian and renders the two unpaired electrons indistinguishable. In this regime, product spin states such as $|\alpha\beta\rangle$ are inadmissible because they do not possess a definite symmetry under particle exchange and therefore violate the Pauli principle (see Appendix). The appropriate eigenstates are instead the singlet and triplet states, which diagonalize the Hamiltonian and have well-defined fermionic symmetry, antisymmetric for the singlet state and symmetric for the triplet state:

$$|S\rangle = \frac{1}{\sqrt{2}}(|\alpha_1\beta_2\rangle - |\beta_1\alpha_2\rangle) \tag{6.14}$$

$$|T_-\rangle = |\beta_1\beta_2\rangle; \quad |T_0\rangle = \frac{1}{\sqrt{2}}(|\alpha_1\beta_2\rangle + |\beta_1\alpha_2\rangle); \quad |T_+\rangle = |\alpha_1\alpha_2\rangle \tag{6.15}$$

$$\tag{6.16}$$

Using the singlet and triplet spin states as a basis for the two-electron spin manifold, the Hamiltonian can be expressed in matrix form, with all matrix elements given in angular frequency units, as

$$\hat{\mathcal{H}} = \begin{array}{c} \\ \langle T_+| \\ \langle S| \\ \langle T_0| \\ \langle T_-| \end{array} \begin{array}{c} \begin{array}{cccc} |T_+\rangle & |S\rangle & |T_0\rangle & |T_-\rangle \end{array} \\ \begin{pmatrix} \omega_{av} - J(r) & 0 & 0 & 0 \\ 0 & J(r) & \omega_{ST_0} & 0 \\ 0 & \omega_{ST_0} & -J(r) & 0 \\ 0 & 0 & 0 & -\omega_{av} - J(r) \end{pmatrix} \end{array} \tag{6.17}$$

where $\omega_{av} = 1/2(\omega_1 + \omega_2)$ and $\omega_{ST_0} = 1/2(\omega_1 - \omega_2)$, and ω_1 and ω_2 are defined as function of the nuclei magnetic quantum numbers m:

$$\omega_1 = g_1 B_0 \frac{\mu_B}{\hbar} + \sum_i a_i m_i, \tag{6.18}$$

$$\omega_2 = g_2 B_0 \frac{\mu_B}{\hbar} + \sum_j a_i m_j \tag{6.19}$$

During the diffusive separation of the radical pair, the electronic spin state evolves coherently and may be conveniently represented within the reduced two-level subspace spanned by the singlet $|S\rangle$ and the central triplet $|T_0\rangle$ states. Accordingly, at any given time the spin wavefunction can be written as a linear superposition of $|S\rangle$ and $|T_0\rangle$:

$$|X\rangle(t) = c_S(t)|S\rangle + c_{T_0}(t)|T_0\rangle, \tag{6.20}$$

which enables the possibility of intersystem crossing. Suppose that a radical pair is initially generated, for example, in the $|T_0\rangle$ spin state and subsequently undergoes diffusive separation. During this interval, the electronic spin state evolves under the influence of the spin Hamiltonian. If, at the time of reencounter, the singlet coefficient c_S differs from zero, the spin wavefunction contains a finite singlet contribution. Upon reencounter, the exchange interaction becomes effective again and enforces projection of the spin state onto the eigenstates of the coupled radical pair. Consequently, within an ensemble of radical pairs, a nonzero fraction will recombine via the singlet channel, indicating that intersystem crossing has occurred during the diffusive evolution.

Such spin-state conversion is facilitated by the distance-dependent structure of the spin energy levels: for sufficiently large electron–electron separations, the singlet $|S\rangle$ and central triplet $|T_0\rangle$ states become degenerate, enabling efficient coherent mixing. In contrast, at a characteristic separation a level crossing may occur between $|S\rangle$ and one of the outer triplet components, $|T_-\rangle$ or $|T_+\rangle$, depending on the sign of the exchange interaction.

In the presence of a strong external magnetic field, however, the Zeeman interaction lifts the degeneracy of the triplet manifold, shifting the T_+ and T_- states far away in energy from the singlet. For fields typically exceeding $\sim 0.1\,\mathrm{T}$, this energy separation is sufficiently large that intersystem crossing involving $|T_+\rangle$ or $|T_-\rangle$ is effectively suppressed. Assuming a sufficiently large spatial separation between the two unpaired electrons, coherent spin-state interconversion is therefore restricted to the $\{|S\rangle, |T_0\rangle\}$ subspace. This singlet–triplet mixing between $|S\rangle$ and $|T_0\rangle$ provides the dominant microscopic pathway for intersystem crossing and underlies the generation of CIDNP in spin-correlated radical pairs under high-field conditions. At low magnetic field the crossing by $|S\rangle \leftrightarrow |T_{\pm 1}\rangle$ becomes possible.

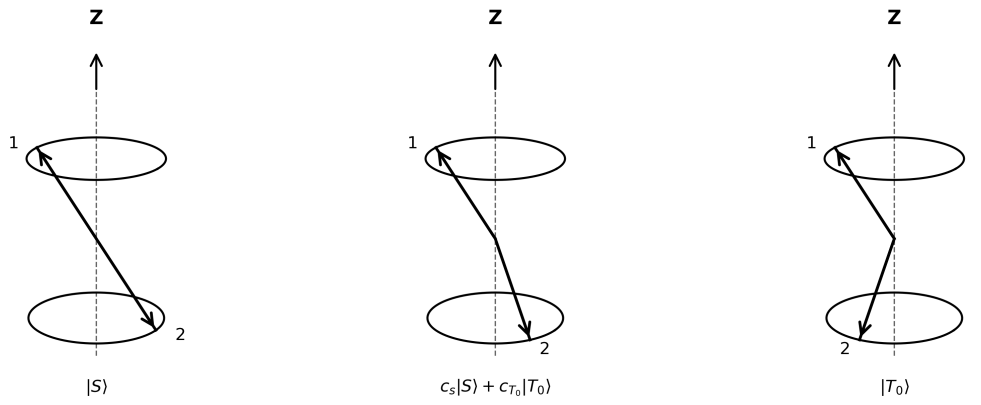


Figure 6.2: Geometric configuration showing two parallel circular trajectories at $z = \pm 1$ and the corresponding vectors connecting the planes.

6.4.3 Vector model

The intersystem crossing mechanism between the singlet S and the T_0 triplet sublevel can be explained in a straightforward and intuitive manner with a vector representation (Fig. 6.2).In these diagrams, the electron spin associated with each radical is depicted as a vector arrow. While such a construction is formally appropriate for two distinguishable, non interacting electron spins, it does not strictly respect the Pauli principle when applied to the correlated singlet and T_0 states. Nevertheless, despite this conceptual inconsistency, the vector model captures the essential physics and yields correct qualitative conclusions.

In the singlet state the total magnetic moment vanishes, which is represented by two antiparallel spin vectors of equal magnitude. By contrast, the T_0 state is characterized by a finite magnetic moment whose projection along the external field (z axis) is zero, this condition is reflected by the relative orientation of the two vectors shown on the right hand side of Fig. 6.2.

If the two radicals do not interact, their electron spins evolve independently in time precessing at different angular frequencies ω_1 and ω_2 (see Eqs. 6.18 and 6.19), unless the radicals are strictly identical, both in their chemical structure and in the complete configuration of their coupled nuclear spins. Consequently, when the radical pair is initially created in the singlet state, the two spin vectors progressively lose their antiparallel alignment. This dephasing will eventually lead to the formation of the T_0 state, then continued evolution will drive the system back towards the singlet configuration, and the process repeats. The interconversion between S and T_0 therefore corresponds to a coherent oscillatory dynamics originating from the difference in electron spin precession frequencies of the two non interacting radicals.

In the high-field regime, where electron and nuclear spin states are separable, the $|S\rangle \leftrightarrow |T_0\rangle$ interconversion is described by a well-defined angular frequency ω_{ST_0}, which characterizes the coherent oscillatory mixing between the two spin states and can be expressed explicitly as follows:

$$\omega_{ST_0} = \frac{1}{2}(\omega_1 - \omega_2) = \frac{1}{2}\Big((g_1 - g_2)B_0\frac{\mu_B}{\hbar} + \sum_i a_i m_i - \sum_j a_j m_j\Big). \tag{6.21}$$

6.4.4 Kaptein's rule

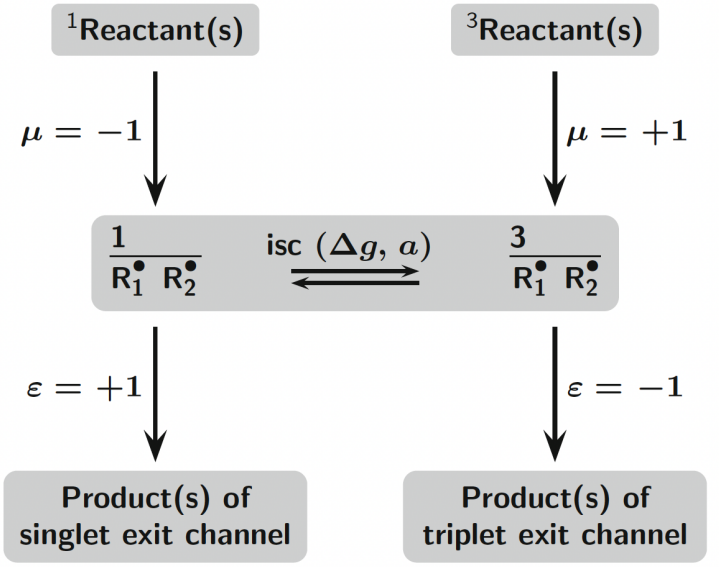


Figure 6.3: From [120]. Radical pairs form with the spin multiplicity inherited from their precursors: ${}^1\overline{(R_1^\bullet R_2^\bullet)}$ from singlet precursors or ${}^3\overline{(R_1^\bullet R_2^\bullet)}$ from triplet precursors. Singlet pairs recombine to form singlet products, whereas triplet pairs usually dissociate into uncorrelated free radicals. Intersystem crossing (isc), driven by g values differences Δg and hyperfine coupling constants a, couples the singlet and triplet manifolds, enabling diagonal pathways such as triplet precursors yielding singlet products. We define a parameter μ for the entry channels (+1 for triplet, −1 for singlet) and a parameter ϵ for the exit channels (with reversed assignments). Their product $\mu \cdot \epsilon = +1$ identifies diagonal pathways, while $\mu \cdot \epsilon = -1$ marks vertical pathways.

It is possible to derive simple qualitative rules [121] that relate the sign of the observed nuclear polarizations (positive for absorption and negative for) to the underlying reaction mechanism and to the magnetic parameters of the radical pairs. The first step is considering that, as already mentioned when describing the diagram in Fig. 6.3, there are two possible pathways from radical pair generation to recombination: vertical or diagonal. This situation can be characterized by introducing a parameter μ for the entry channel, with $\mu = +1$ denoting a triplet precursor and $\mu = -1$ a singlet precursor, together with a parameter ϵ for the exit channel. The latter takes the value $\epsilon = +1$ when product formation proceeds from singlet radical pairs and $\epsilon = -1$ when it originates from triplet radical pairs, including cases in which the observed product is formed via intermediate free radicals. The product $\mu\epsilon$ is then equal to +1 for pathways that follow a diagonal connection in the scheme, and to −1 for pathways corresponding to vertical transitions.

To set the connection between the magnetic parameters and the sign of the signal amplitudes it is convenient to switch to a rotating frame coordinate system with the average frequency:

$$\omega_{av} = \frac{1}{2}(\omega_1 + \omega_2) = \frac{1}{2}\Big((g_1 + g_2)B_0\frac{\mu_B}{\hbar} + \sum_i a_i m_i + \sum_j a_j m_j\Big). \tag{6.22}$$

If we consider a radical pair with only one observed nucleus of spin 1/2, generated in a triplet state, we see that In this system, the angular frequencies of the radicals depend

solely on the $\Delta g = g_1 - g_2$ and on the nuclear spin state: the Zeeman and hyperfine contributions reinforce each other for nuclear spin state α but partially cancel for state β. As a consequence, singlet-products are enriched in α sstate and depleted in β state, resulting in an enhanced absorptive NMR signal. An emissive signal is obtained if the sign of either Δg or the hyperfine coupling constant a_i is inverted, but not if both are inverted simultaneously.

These observations can be formalized in the Kaptein's rule for the phase Γ_i of a CIDNP net effect:

$$\Gamma(i) = \mu \times \epsilon \times sign(\Delta g) \times sign(a_i). \tag{6.23}$$

Similar argument can be used to derive the Kaptein's rule for a CIDNP multiplet effect Γ_{ij} of two observed nuclei i and j:

$$\Gamma_{ij} = \mu \times \epsilon \times sign(a_i) \times sign(a_j) \times \sigma_{ij} \times sign(J_{ij}), \tag{6.24}$$

where σ_{ij} is $+1$ when two nuclei are contained in the same radical and -1 when they are not, and J_{ij} is the coupling constant between the nuclei. Γ_{ij} is $+1$ when the sequence is emission/absorption (A/E) and -1 when the sequence is absorption/emission (A/E).

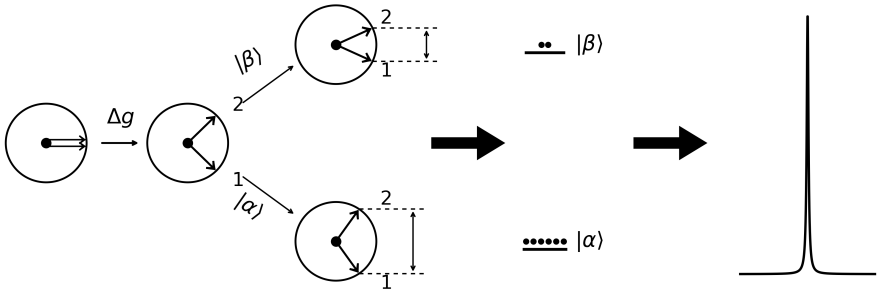


Figure 6.4: Using the vector model (projection on the xy plane), this diagram illustrates the spin evolution of a triplet-born radical pair with one spin-$\frac{1}{2}$ nucleus (radical 1). The relative signs of the Δg and hyperfine coupling determine whether Zeeman and hyperfine contributions add or cancel, thereby fixing the CIDNP phase in accordance with Kaptein's rule.

6.5 Photo-CIDNP for Fragment Screening

Fragment-based drug discovery (FBDD) relies on the detection of weak binders, typically small molecules below 300 Da with affinities in the μM to mM range, against a protein target. While NMR is a reference method for this purpose, its inherent low sensitivity demands high sample concentrations and long acquisition times. Photo-chemically induced dynamic nuclear polarization (photo-CIDNP) addresses this limitation by enhancing the ligand signal by one to two orders of magnitude, enabling detection at low micromolar concentrations in a single-scan experiment of a few seconds duration [11, 12].

To enable the detection of ligand signals using photo-CIDNP NMR, the ligand must participate in a photoinduced radical-pair process. This is typically achieved by introducing a photosensitizer (PS), a molecular light absorber that is able to modify the course

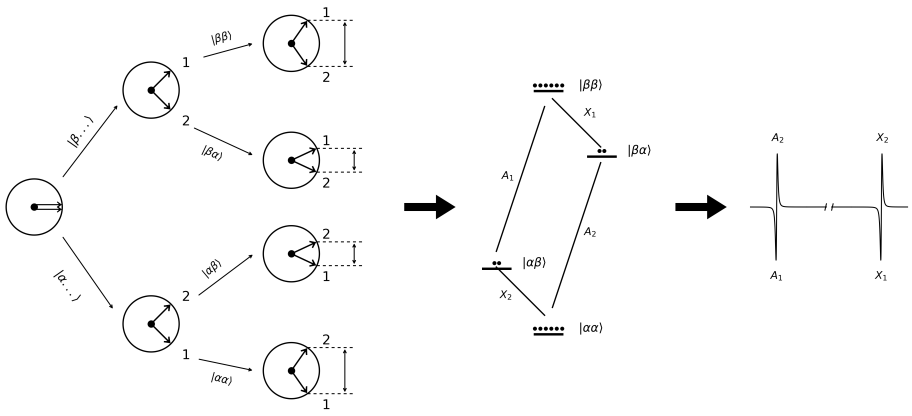


Figure 6.5: Using the vector model (projection onto the xy plane), this diagram illustrates the simplest case of a triplet-born radical pair containing two spin-$\frac{1}{2}$ nuclei (A and X) in the same radical, both with positive hyperfine couplings and singlet product formation. Setting $\Delta g = 0$ simplifies the analysis and yields an E/A multiplet, with emission of the low-field and absorption of the high-field component of each doublet, assuming $a_A > a_X$.

of a photochemical reaction. Upon absorption of a photon, the photosensitizer is promoted to an electronically excited singlet state and, subsequently undergoes intersystem crossing to a longer-lived triplet state. From this excited configuration, it can engage in transient electron- or energy-transfer interactions with a nearby ligand (L), resulting in the formation of a radical pair with initially correlated electron spins:

$$PS + L \leftrightarrow L^* + PS^*, \tag{6.25}$$

where the asterisk indicates the hyperpolarised states. These spin correlations provide the necessary starting condition for spin-selective reaction and recombination pathways, ultimately leading to the generation of non-Boltzmann nuclear spin populations and to the enhanced NMR signal intensities characteristic of photo-CIDNP experiments.

When these radical pairs are generated in the presence of an external magnetic field, their spin-selective recombination pathways are influenced by the nuclear spin configuration. Although the chemical identity of the recombination products is indistinguishable from that of the initial reactants, the associated nuclear spin populations are redistributed in a manner that deviates from thermal (Boltzmann) equilibrium.

Eq. 6.25 describes a second-order reaction between the photosensitizer and the ligand, since the reaction rate depends on the concentration of both reactants and can be described by the kinetic law rate $= k[PS][L]$. However, because the concentration of the ligand is typically kept in large excess compared to that of the photosensitizer, the reaction can be treated as pseudo-first-order with respect to the photosensitizer concentration:

$$L \underset{R_1}{\overset{k_{cidnp}}{\rightleftharpoons}} L^*, \tag{6.26}$$

where k_{cidnp} is the rate associated with radical pair formation within the photo-CIDNP process, and R_1 rate associated with the back relaxation of the hyperpolarised ligand to thermal equilibrium.

When considering the presence of a protein (P) which constitutes the binding partner

of the ligand, the reaction equation becomes:

$$PL \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} L \underset{R_{1,L^*}}{\overset{k_{cidnp}}{\rightleftharpoons}} L^* \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL^* \xrightarrow{R_{1,PL^*}} PL. \tag{6.27}$$

A fraction of ligand molecules will bind with the protein forming the intermolecular protein-ligand complex (PL). Therefore, there will be less free ligands available to produce radical pairs with the photosensitizer. Moreover, the relaxation rate of the bound hyperpolarised ligand is higher with respect to the relaxation rate of the free hyperpolarised ligand ($R_{1,PL^*} > R_{1,L^*}$). This increase is specific to non-equilibrium polarization. For nuclear spin systems at thermal equilibrium, longitudinal relaxation is governed by the spectral density function $J(\omega)$, whose value at the relevant Larmor frequencies decreases with increasing rotational correlation time, that is with slower molecular tumbling. The expression for the spectral density function is the following [122]:

$$J(\omega) = \frac{2}{5}\frac{S^2\tau_c}{1+(\tau_c\omega)^2} + \frac{2}{5}\frac{(1-S)^2\left(\frac{\tau_c\tau_e}{\tau_c+\tau_e}\right)}{1+\left(\frac{\tau_c\tau_e}{\tau_c+\tau_e}\omega\right)^2} \tag{6.28}$$

where τ_c and τ_e are respectively the rotational and internal correlation times and S is the order parameter describing the amplitude of internal motions.

By contrast, in non-Boltzmann spin states generated by hyperpolarization, an additional contribution to selective longitudinal relaxation arises from the autorelaxation term of the bound hyperpolarised ligand. This contribution is proportional to the zero-frequency spectral density $J(0)$, which scales linearly with τ_c, and thus increases with increasing tumbling time. This qualitative difference underlies the enhanced sensitivity of hyperpolarised NMR experiments to binding-induced changes in molecular dynamics.

In fragment-based drug discovery (FBDD), the weak binding affinity of fragment ligands results in dissociation rates k_{off} that are significantly higher than the rate at which the ligand relaxes from the hyperpolarised state back to thermal equilibrium in the bound form R_{1,PL^*}:

$$k_{off} >> R_{1,PL^*}. \tag{6.29}$$

NMR-based fragment screening typically targets weak to moderate affinity ligands. Binding affinity describes the strength of the interaction between a biomolecule, such as a protein or nucleic acid, and a ligand or binding partner. This interaction strength is most commonly quantified by the equilibrium dissociation constant K_D, which provides a consistent basis for evaluating and comparing bimolecular binding interactions. A reversible binding reaction can be represented in the kinetic form

$$P + L \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} PL, \tag{6.30}$$

where k_{on} and k_{off} are the association and dissociation rate constants, respectively. At equilibrium, the forward and reverse rates are equal:

$$k_{\text{on}}[P][L] = k_{\text{off}}[PL], \tag{6.31}$$

and the equilibrium dissociation constant can be defined as

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{\text{off}}}{k_{\text{on}}}. \tag{6.32}$$

Typical dissociation constants values range from the high μM to low mM range for weak to moderate affinity ligands. Therefore, assuming typical on-rates k_{on} on the order of the diffusion limit ($10^9\,\text{M}^{-1}\text{s}^{-1}$), and dissociation constants K_D in the mM to μM range typical of FBDD fragments, the corresponding off-rates k_{off} are in the range of 10^3 to $10^6\,\text{s}^{-1}$, corresponding to residency times of μs to ms. Since typical longitudinal relaxation times are in the range of s to high ms, the condition $k_{\text{off}} \gg R_{1,PL^*}$ is well satisfied in most cases. Therefore, in this case the equation reduces to:

$$PL \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} L \underset{R_{1,eff}}{\overset{k_{cidnp}}{\rightleftharpoons}} L^*, \tag{6.33}$$

where $R_{1,eff} = pR_{1,L^*} + (1-p)R_{1,PL^*}$ is the effective relaxation rate of the hyperpolarised ligand, which takes into account the contribution from both the free and bound forms, and $p = \frac{L}{L+PL}$ is the fraction of free ligand molecules. The exchange equilibrium $\text{PL}^* \rightleftharpoons \text{L}^*$ is reached on a timescale much shorter than that of photo-CIDNP polarization and is therefore assumed to hold at all times. From Eq. 6.33, we can write the equation for the time evolution of the hyperpolarised ligand due to photo-CIDNP effect:

$$\frac{dL^*}{dt} = k_{cidnp}L - R_{1,eff}L^*. \tag{6.34}$$

Since typical photo-CIDNP signal enhancements of approximately $20-50$-fold correspond to ligand polarizations in the range of $0.8-2.0\,\%$, the resulting hyperpolarised ligand population (L^*) represents only a small fraction of the total ligand pool. Under these conditions, the concentration of thermally polarised ligand L can be treated as effectively constant over the timescale of the experiment. With this assumption, the time evolution of the hyperpolarised ligand population can be solved analytically, yielding:

$$L^*(t) = \frac{k_{cidnp}}{k_{cidnp} + R_{1,eff}} L_0(1 - e^{-(k_{cidnp}+R_{1,eff})t}), \tag{6.35}$$

where $L_0 = L_{t=0} = L + L^*$ correspond to the initial concentration of the free ligand.

Two limiting regimes can be distinguished for the time evolution of the hyperpolarised ligand concentration. At short times, a Taylor expansion of the solution yields

$$L^*(t) \xrightarrow{t\to 0} k_{\text{cidnp}} L_0\, t - \frac{k_{\text{cidnp}}(k_{\text{cidnp}} + R_{1,\text{eff}})\, L_0}{2}\, t^2 + \ldots, \tag{6.36}$$

indicating an initial linear build-up governed by the photo-CIDNP rate, followed by curvature arising from relaxation processes. In the long-time limit, that is for time larger than $1/(k_{cidnp} + R_{1,eff})$,the system approaches a stationary state,

$$L^*(t) \xrightarrow{t\to\infty} \frac{k_{\text{cidnp}}}{k_{\text{cidnp}} + R_{1,\text{eff}}}\, L_0, \tag{6.37}$$

which reflects the balance between continuous polarization transfer and effective longitudinal relaxation.

The simultaneous reduction of the free ligand concentration and the increase of the effective longitudinal relaxation rate upon binding result in an attenuation of the photo-CIDNP hyperpolarised signal in the presence of a protein (see Figure 6.6).

Experimentally, this manifests as a decrease in the corresponding NMR peak integrals. Binding events can therefore be detected by introducing a normalized polarization ratio

(PR), defined to lie between 0 and 1, where values close to unity indicate the absence of interaction and values approaching zero correspond to complete binding of the ligand population:

$$\text{Polarization Ratio} = \frac{\Gamma_{PL}}{\Gamma_L}. \tag{6.38}$$

The peak integrals Γ_L and Γ_{PL} are obtained from experiments recorded in the absence and presence of protein, respectively. In the short-time regime, as follows from the Taylor expansion of the polarization build-up, $L^*(t) \xrightarrow{t\to 0} k_{\text{cidnp}} L_0\, t + \mathcal{O}(t^2)$, relaxation effects are negligible and the photo-CIDNP signal is proportional to the concentration of free ligand. The polarization ratio therefore depends only on the free ligand concentration in the presence and absence of protein. Since this concentration is governed by the binding equilibrium, the polarization ratio can be expressed in terms of the dissociation constant K_D and the total ligand and protein concentrations, L_{tot} and P_{tot}.

For the equilibrium reaction $\mathrm{P} + \mathrm{L} \rightleftharpoons \mathrm{PL}$, the dissociation constant is defined as

$$K_D = \frac{[\mathrm{P}][\mathrm{L}]}{[\mathrm{PL}]} = \frac{(L_{\text{tot}} - PL)\,(P_{\text{tot}} - PL)}{PL}, \tag{6.39}$$

where in the last term we used the mass conservation $L_{\text{tot}} = L + PL$ and $P_{\text{tot}} = P + PL$. Defining the polarization ratio as the normalized bound fraction $x = PL/L_{\text{tot}}$ yields

$$K_D = \frac{(P_{\text{tot}} - x)(L_{\text{tot}} - x)}{x}, \tag{6.40}$$

which leads to the quadratic equation

$$L_{\text{tot}} x^2 - (P_{\text{tot}} + L_{\text{tot}} + K_D)x + P_{\text{tot}} L_{\text{tot}} = 0. \tag{6.41}$$

Solving for x provides an explicit expression for the polarization ratio as a function of the dissociation constant,

$$\text{PR} = \frac{L_{\text{tot}} + P_{\text{tot}} + K_D + \sqrt{(P_{\text{tot}} + L_{\text{tot}} + K_D)^2 - 4P_{\text{tot}}L_{\text{tot}}}}{2L_{\text{tot}}}, \tag{6.42}$$

which can be used to extract K_D from experimental photo-CIDNP data.

At longer times, relaxation effects become increasingly important and further reduce the polarization ratio, thereby enhancing the contrast between experiments performed with and without protein. In practical photo-CIDNP screening experiments, sufficient signal-to-noise at low micromolar concentrations typically requires illumination times of several hundred milliseconds. The measurements presented in this work were therefore carried out in the steady-state regime, which is reached for times exceeding $(R_{1,\text{eff}} + k_{\text{cidnp}})^{-1}$ and typically lies in the 1–2 s range.

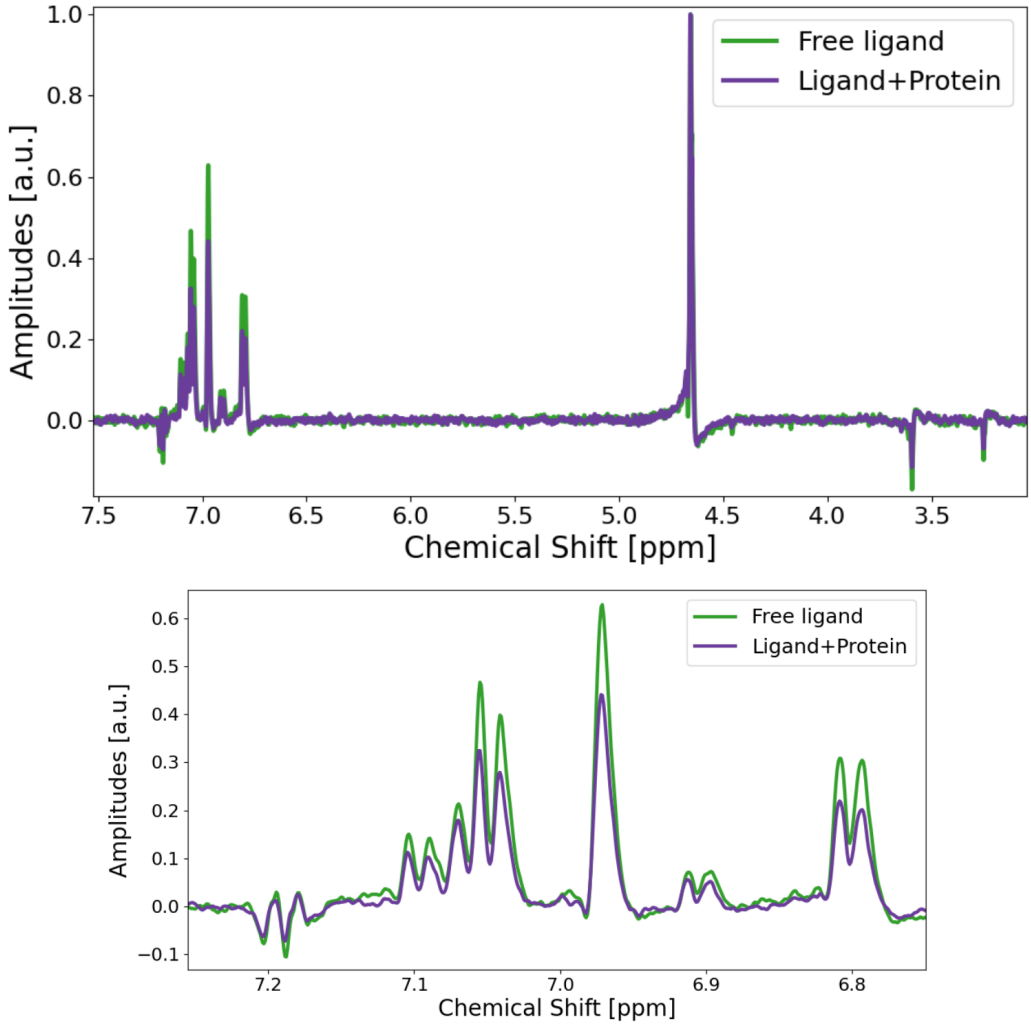


Figure 6.6: Comparison between the photo-CIDNP ^{1}H NMR spectra of the ligand in isolation and in the presence of the protein of interest for one of the small molecules in the screening dataset (see Section 7.2.1). The upper panel shows the full spectral region, while the lower panel displays a magnified view (inset) highlighting the variations in signal intensity between the ligand-only and ligand+protein conditions. A clear quenching of the CIDNP-enhanced signals is observed in the presence of the protein, indicating ligand binding.

Chapter 7

Detecting ligand-protein binding

7.1 The pipeline

We developed an analytical pipeline combining deep learning and rule-based approaches to automatically screen relevant binding candidates using ^{1}H NMR spectra. The pipeline operates through sequential steps, correlating information across different spectral modalities to characterize and assign the observed signals.

7.1.1 The input data

The framework is designed to process two inputs: spectral data and molecular structure representations. To determine the binding interaction between a ligand and a protein using photo-CIDNP NMR spectroscopy, it is necessary to acquire at least four spectra under different experimental conditions. First, photo-CIDNP spectra (hereafter referred to as the *light spectrum*, since the sample is illuminated) of the ligand in isolation and of the ligand in the presence of the protein of interest must be recorded. In typical fragment screening experiments the protein concentration is kept sufficiently low that the protein resonances contribute negligibly to the ^{1}H spectrum. As a consequence, ligand signals can generally be analysed without the need for dedicated suppression of protein resonances. In addition, to identify photo-CIDNP-active signals, spectra obtained in the absence of illumination, i.e. standard ^{1}H NMR spectra (hereafter referred to as the *dark spectrum*), must also be acquired. By comparing the amplitudes of corresponding resonances in the light and dark spectra, hyperpolarised signals can be systematically identified. This amplitude-based detection enables a direct comparison between ligand-only spectra and spectra acquired in the presence of the protein, thereby providing a consistent criterion for assessing photo-CIDNP enhancements and their modulation upon ligand–protein interaction. As a result, the minimal dataset required for analysis comprises four spectra: dark and light spectra for the free ligand, and dark and light spectra for the ligand in the presence of the protein.

Further experimental flexibility can be introduced by varying the T_2 relaxation delays in order to monitor the evolution of signal intensities (see Fig. 7.1). Extended relaxation delays increase the sensitivity to binding by enhancing the contrast between the ligand-only spectrum and the spectrum recorded in the presence of the protein, that is the signal quenching upon binding is larger. However, longer delays also lead to a reduction in overall signal-to-noise ratio, making it necessary to find a suitable compromise between

sensitivity to the binding effect and spectral quality. In this context, the developed framework has been designed to accommodate multiple spectra recorded under a range of relaxation delays, automatically evaluating the relative noise levels and signal intensities to determine the most favorable acquisition parameters for downstream quantification of binding interactions.

One of the key strengths of photo-CIDNP NMR spectroscopy is its ability to provide atom-level information on binding strength, enabling the generation of an epitope map of the ligand. Automating this part of the analysis requires assigning each hydrogen atom in the molecule to its corresponding NMR signal, which in turn requires knowing the structure of the ligand. This structural information is supplied to the framework as an SD file in the V2000 molfile format. Each record includes a header and a counts line indicating the total number of atoms and bonds, followed by an atom block listing all atoms with their 2D Cartesian coordinates, element types, and optional properties such as charge or isotope information. A bond block then specifies molecular connectivity, including bonded atoms, bond orders, and stereochemical annotations.

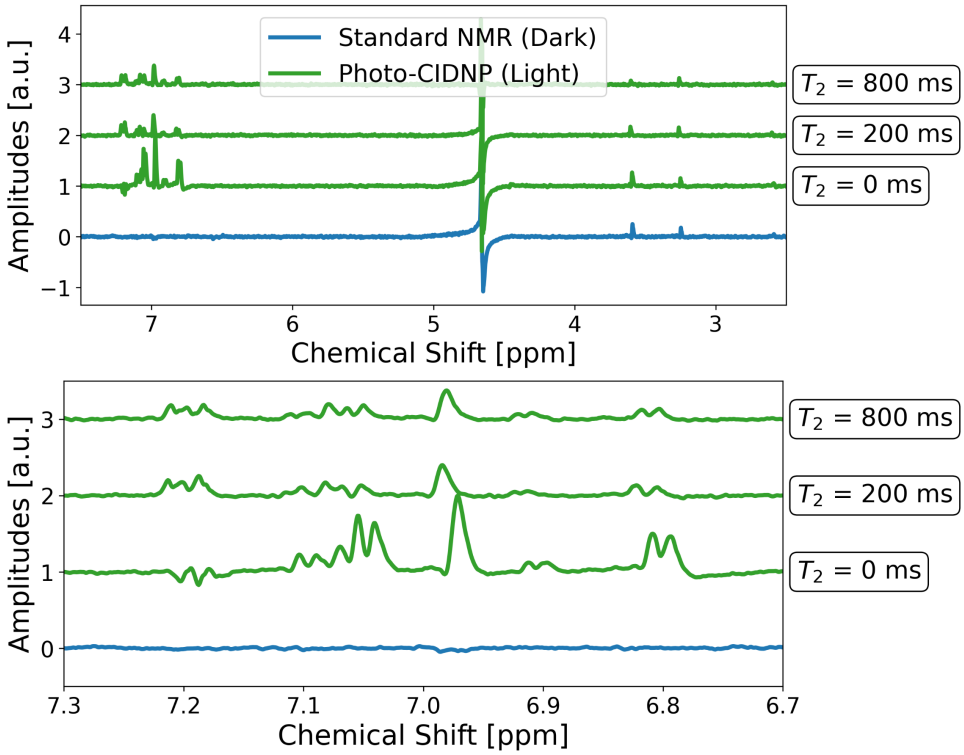


Figure 7.1: Comparison between the standard ^{1}H NMR spectrum and three photo-CIDNP spectra acquired with increasing T_2 delays for the same ligand. The top panel shows the full spectral region, while the bottom panel displays an inset highlighting a selected portion of the spectrum.

7.1.2 Spectral pre-processing

The spectral data require preprocessing prior to automated analysis for quantifying ligand-protein interactions. The spectral files, provided in Bruker format[1], are read to

[1]At present, this is the only file format supported by the algorithm. Future versions of the software are planned to include support for additional file formats.

extract a dictionary containing both acquisition and processing parameters, as well as the raw free induction decay (FID) signal. The first preprocessing step involves correcting for the digital filter. While Bruker spectrometers apply digital filtering to suppress high-frequency noise and prevent aliasing, the filtering process introduces a group delay at the start of the Free Induction Decay (FID). This delay causes severe first-order phase distortions, making its correction or removal essential for accurate downstream analysis. Subsequently, the number of data points is doubled via zero-filling, a procedure that enhances the digital resolution in the frequency domain without altering the intrinsic spectral information. To further improve the signal-to-noise ratio prior to Fourier transformation, the FID is multiplied by a decaying exponential function, a line-broadening technique, with a time constant chosen to be inversely proportional to the spectral width.

The correct identification of signals produced by the same ligand atom across different spectra (recorded at different T_2 relaxation delays and for the ligand alone or in the presence of a protein) is crucial for an accurate quantification of binding strength. In this respect, it is fundamental to apply the same phase correction to all spectra of a ligand–protein pair. The phase correction can be optimised automatically. Since the distinctive trait of an unphased spectrum is the presence of peaks with a dispersive lineshape in the real spectrum, the algorithms for automatic phasing minimise the occurrence of signals with negative amplitudes. However, this approach cannot be applied to photo-CIDNP NMR data, since hyperpolarisation can lead to absorption-mode peaks with negative amplitudes.

We can nevertheless exploit two additional properties of correctly phased NMR spectra. Compared with the initial Fourier-transformed data, a properly phased spectrum exhibits spectral features that are smoother and structurally simpler. In general, the simplest representation of a signal corresponds to a form in which its information entropy is minimized; this motivates the use of a Shannon-type entropy measure as an objective criterion for phase optimization [76].

Moreover, for a correctly phased spectrum, the real part should contain only absorption-mode contributions, while the imaginary part should contain only dispersion-mode contributions [123]. Owing to the odd parity of the dispersion lineshape with respect to the resonance frequency, the integral of the imaginary part vanishes when evaluated over a spectral window that is symmetric about the resonance. In contrast, an unphased spectrum corresponds to a mixture of absorption and dispersion components in both the real and imaginary channels. This situation can be illustrated explicitly as

$$
\begin{aligned}
&\textbf{Spectrum with phase distortion } \phi(\Omega):\\
&S(\Omega) = S_0\Big(\cos\phi\, A(\Omega) - \sin\phi\, D(\Omega)\Big) + iS_0\Big(\cos\phi\, D(\Omega) + \sin\phi\, A(\Omega)\Big),\\
&\int_{\text{ppm range}} \operatorname{Im} S(\Omega)\,\mathrm{d}\Omega = \int_{\text{ppm range}} S_0\Big(\cos\phi\, D(\Omega) + \sin\phi\, A(\Omega)\Big)\,\mathrm{d}\Omega \neq 0,\\
&\textbf{Phase-corrected spectrum:}\\
&S(\Omega) = S_0\Big(A(\Omega) + iD(\Omega)\Big),\\
&\int_{\text{ppm range}} \operatorname{Im} S(\Omega)\,\mathrm{d}\Omega = \int_{\text{ppm range}} S_0 D(\Omega)\,\mathrm{d}\Omega = 0.
\end{aligned}
\tag{7.1}
$$

It is important to emphasize that neither of these conditions imposes any constraint on the sign of the spectral peaks; in particular, both absorptive and emissive signals are compatible with the above criteria.

On this basis, the automatic phase correction procedure proposed here builds upon the numerical strategy originally introduced by Ernst [123] and the ACME approach [76]. The optimal phase parameters are obtained by minimizing a cost function with respect to the zero- and first-order phase angles ϕ_0 and ϕ_1, composed of two contributions: (i) a Shannon entropy term computed from the first derivative of the real part of the spectrum, and (ii) a term enforcing the vanishing integral of the imaginary component. The phased real and imaginary parts are obtained as

$$S(\phi_0, \phi_1) = \tilde{S}\, e^{i(\phi_0 + \phi_1 k)}, \tag{7.2}$$

where $\tilde{S}$ is the complex spectrum and k is the point index. The objective function is then defined as

$$\begin{aligned} \min_{\phi_0,\,\phi_1} E &= -\sum_i h_i \ln h_i + P, \\ h_i &= \frac{|R_i^{(1)}|}{\sum_j |R_j^{(1)}|}, \\ P &= \gamma \left| \sum_j I_j \right|, \end{aligned} \tag{7.3}$$

where $R_i^{(1)}$ denotes the first derivative of the real part $R_i = \mathrm{Re}[S_i(\phi_0, \phi_1)]$, $I_j = \mathrm{Im}[S_j(\phi_0, \phi_1)]$ denotes the imaginary component, and γ is a weighting factor.

However, in some cases, finding the optimal phase correction automatically is challenging. When the buffer used for sample preparation introduces predominant signals, phase optimisation algorithms tend to phase these dominant resonances correctly while leaving the weak CIDNP signals poorly phased. Furthermore, the strong buffer peaks can introduce baseline distortions that additionally complicate automatic phase correction. In such cases, the algorithm allows manual phase correction. Phase distortions in one-dimensional NMR spectra arise from several experimental factors that contribute differently to zero- and first-order phase errors [76]. Zero-order phase misadjustments originate primarily from a mismatch between the reference phase and the phase of the receiver detection chain. In contrast, first-order phase errors are associated with time delays between radiofrequency excitation and signal acquisition, as well as with frequency-dependent effects such as variations in the effective flip angle across the spectral bandwidth and phase shifts introduced by analog or digital filtering used to suppress out-of-band noise. Since both zero- and first-order phase distortions are determined by fixed properties of the experimental setup and acquisition protocol, they remain largely constant across spectra recorded under identical conditions. Consequently, an optimal phase correction can be obtained by manually phasing a single representative spectrum. The resulting phase parameters may then be consistently applied to all spectra acquired within the same screening project, provided that acquisition and processing settings are unchanged.

An alternative strategy to mitigate phasing-related issues is the use of magnitude spectrum (see Figure 7.2). Magnitude spectra, constructed by taking the modulus of the complex Fourier-transformed NMR signal, are inherently insensitive to phase errors:

$$S_M(\Omega) = \sqrt{Re(S(\Omega))^2 + Im(S(\Omega))^2} \tag{7.4}$$

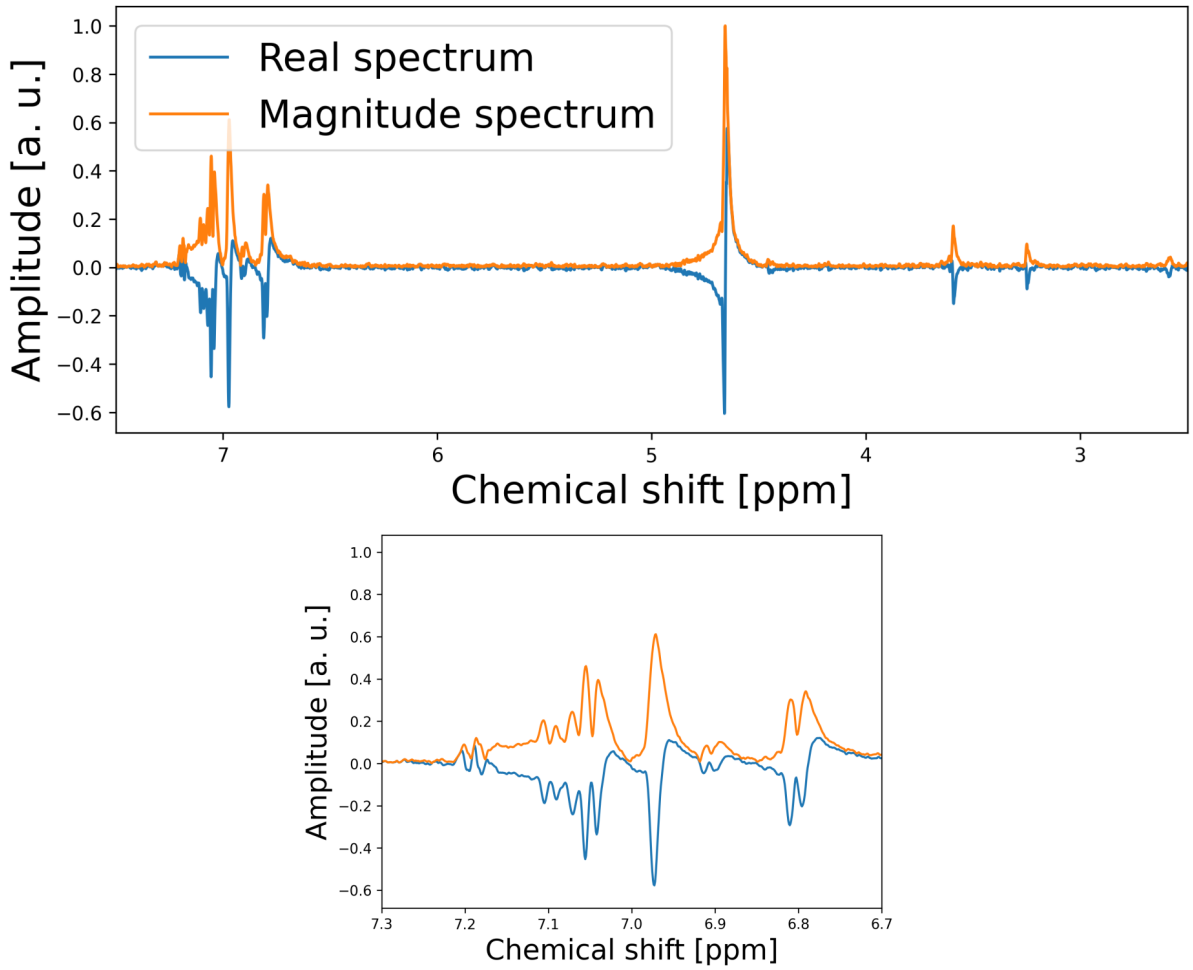


Figure 7.2: Comparison between the real (phase-sensitive) and magnitude representations of a ^{1}H NMR spectrum for one of the samples in the screening dataset (see Section 7.2.1). This representation removes phase distortions that may affect the real spectrum and facilitates robust downstream spectral analysis.

In contrast to conventionally phased absorption spectra, where imperfect zero-order or first-order phasing can lead to distorted line shapes, asymmetric peaks, or biased integrals, magnitude spectra eliminate the need for explicit phase correction. As a result, all resonances appear strictly positive and retain approximately symmetric line shapes, independent of residual acquisition or processing imperfections. This property is particularly advantageous in automated processing pipelines and high-throughput analyses, where manual phasing is impractical or inconsistent. Moreover, in pattern-recognition applications [69], such as machine-learning-based spectral classification or database matching, phase invariance ensures uniform treatment of signals and prevents sign inversions or phase-dependent distortions, thereby enabling more robust extraction of spectral features based on peak intensities, integrals, or statistical descriptors.

The use of the magnitude spectrum introduces a non-linearity in the signal representation, which modifies the statistical properties of the noise [68]. While the noise in the real part of the spectrum is Gaussian distributed with standard deviation σ, the magnitude operation leads to a Rician distribution of the noise in the magnitude spectrum. In this case, the distribution depends on the ratio $\bar{M}/\sigma$, where M denotes the amplitude in the magnitude spectrum and σ the standard deviation of the Gaussian noise in the real

channel.

The discrepancy between Gaussian and Rician statistics is particularly pronounced at low signal-to-noise ratios, especially for $A/\sigma \leq 1$, where A is the signal amplitude in the real spectrum. In this regime, the magnitude operation introduces a positive bias and a clear deviation from Gaussian behaviour. However, already at $A/\sigma = 3$, the Rician distribution closely approximates a Gaussian distribution, and the bias becomes comparatively small.

In baseline regions, where no true signal is present ($A = 0$), the Rician distribution reduces to a Rayleigh distribution. In this case, the mean noise level in the magnitude spectrum is related to the Gaussian standard deviation in the real channel according to

$$\text{noise}_M = \sigma_{\text{Gaussian}} \sqrt{2 - \frac{\pi}{2}}. \tag{7.5}$$

This transformation also introduces a positive bias in the amplitude of the magnitude signal, which becomes particularly relevant at very low signal-to-noise ratios. Even in the absence of a true signal, the magnitude operation yields a strictly positive expectation value due to the underlying Rayleigh statistics of the noise. As a consequence, weak signals may appear artificially enhanced.

A simple post-processing correction can be applied to partially compensate for this bias. Assuming a Gaussian noise level σ in the real channel, a commonly used approximation for the bias-corrected magnitude amplitude $\hat{A}$ is

$$\hat{A} = \begin{cases} \sqrt{M^2 - \sigma^2}, & M > \sigma, \\ 0, & M \leq \sigma, \end{cases} \tag{7.6}$$

where M denotes the measured magnitude signal. This correction reduces the positive bias introduced by the magnitude operation and provides a more accurate estimate of the underlying true signal amplitude, particularly in the low signal-to-noise regime.

Finally, it is important to note that the linewidth of peaks increases in the magnitude spectrum as a consequence of the non-linear transformation. Even in the absence of noise, the magnitude operation modifies the analytical form of the lineshape. For a Lorentzian signal in the real spectrum, the corresponding magnitude lineshape exhibits a broader full width at half maximum (FWHM). This broadening can be computed analytically: for an ideal Lorentzian lineshape, the FWHM in the magnitude spectrum is increased by a factor of $\sqrt{3}$ compared to the original Lorentzian in the real part. This effect should be taken into account when comparing linewidths or when defining integration windows and peak-picking criteria in the magnitude domain.

Following baseline correction, the dark spectra are aligned to their corresponding light spectra. This alignment step is essential to ensure that signals originating from the same nuclear sites are consistently matched across different experimental conditions. Illumination of the sample, which is required to induce the CIDNP effect, results in a modest temperature increase that causes small but systematic shifts of resonance frequencies relative to standard NMR spectra. These variations are routinely corrected using linear temperature coefficients ($\Delta\delta/\Delta T$), which model how the dynamic averaging of the molecule's conformations and solvent interactions change with thermal energy [124, 125]. To compensate for these shifts, spectral alignment is carried out by computing the cross-correlation between the dark and light spectra over a range of relative lags:

$$(f \star g)[\tau] = \sum_k f[k]\, g[k + \tau], \tag{7.7}$$

where f and g denote the dark and light spectra respectively and τ is the lag index. The lag that maximizes the cross-correlation,

$$\tau^* = \arg\max_{\tau} (f \star g)[\tau], \tag{7.8}$$

is selected, and the dark spectrum is subsequently shifted along the frequency axis to achieve optimal alignment.

7.1.3 Peak picking

After pre-processing, the algorithm proceeds with the identification and characterization of spectral signals that are hyperpolarised by the photo-CIDNP effect, hereafter referred to as *photo-active* signals. This is achieved by detecting both positive and negative peaks in the dark and light spectra, a step that is naturally omitted when processing absolute magnitude spectra. Automated peak picking was performed using the nmrglue Python package [126] to identify signals exceeding an absolute intensity threshold set to six times the baseline noise level. To ensure robust signal selection and exclude baseline artefacts, the candidate peaks were subsequently filtered by their topographic prominence, a measure of a peak's relative height from its local baseline, with a minimum requirement set to three times the noise level. Detected peaks are subsequently matched using the Hungarian assignment algorithm [127].

Peak matching is formulated as a bipartite assignment problem. Let $\{\mathcal{P}_i^{\mathrm{d}}\}_{i=1}^{N_{\mathrm{d}}}$ and $\{\mathcal{P}_j^{\mathrm{l}}\}_{j=1}^{N_{\mathrm{l}}}$ denote the sets of peaks identified in the dark and light spectra, respectively. Each peak $\mathcal{P}$ is characterized by its central frequency ω and by a linewidth-defined interval $\mathcal{I} = [\omega^-, \omega^+]$, referred to as the peak range. The optimal assignment π minimizes the total cost

$$\min_{\pi} \sum_{i} C_{i,\pi(i)}, \tag{7.9}$$

with pairwise costs defined as

$$C_{ij} = \frac{1}{2}\left(1 - \mathrm{IoU}(\mathcal{I}_i^{\mathrm{d}}, \mathcal{I}_j^{\mathrm{l}}) + \frac{|\omega_i^{\mathrm{d}} - \omega_j^{\mathrm{l}}|}{\max_{k,l}|\omega_k^{\mathrm{d}} - \omega_l^{\mathrm{l}}|}\right). \tag{7.10}$$

where $\mathrm{IoU}(\mathcal{I}_i^{\mathrm{d}}, \mathcal{I}_j^{\mathrm{l}})$ denotes the Intersection over Union of the corresponding peak ranges. Peaks are not required to be matched, since the photo-CIDNP effect may render visible signals that are absent in the dark spectrum. Moreover, some detected peaks may not originate from the ligand but instead arise from the solvent, sample matrix, or experimental artefacts.

Following peak matching, photo-activity is determined by comparing peak amplitudes in the dark and light spectra. For a matched pair of peaks, the i-th peak is classified as photo-active if

$$|A_i^{\mathrm{l}}| \geq 2.5\,|A_i^{\mathrm{d}}|, \tag{7.11}$$

where A_i^{l} and A_i^{d} denote the peak amplitudes in the light and dark spectra, respectively. A peak detected in the light spectrum that is not matched to any dark peak is classified as photo-active if

$$|A_j^{\mathrm{l}}| \geq 3\,\sigma, \tag{7.12}$$

where σ denotes the estimated noise level of the spectrum.

The candidate photo-active peaks are subsequently assembled into multiplets using MuSe Net [52, 83], a probabilistic deep learning framework for the automatic classification of splitting patterns (read Chapter 4 and 5). MuSe Net was originally developed for standard ^{1}H NMR spectra, which exclusively contain signals of positive intensity. Nevertheless, the model could be successfully applied to photo-CIDNP spectra, which, owing to hyperfine coupling and g-factor effects, comprise both positive and negative intensity signals. The only adaptation required was to account for the presence of negative peaks. In practice, this was accomplished by inverting the sign of negative peaks, either by taking the absolute value of the real part of the spectrum or by using the magnitude spectrum. Since the output of MuSe Net is a segmentation of the spectral range, the resulting mask can be directly applied to the original spectrum, allowing the locations of spectral features to be recovered irrespective of their sign.

This application highlights the robustness and generalization capacity of MuSe Net beyond its initial design scope. Furthermore, MuSe Net predictions serve not only to assemble individual hyperpolarised peaks into multiplets, but also to suppress false positives. Specifically, only those picked peaks that lie within a region predicted by MuSe Net are retained.

7.1.4 Bayesian signal assignment

We aim to assign the hyperpolarised signals to their corresponding hydrogen atoms. This step is crucial because the measured signal quenching is not uniform across all spectral peaks; rather, it depends on the binding strength, which varies according to the atom's position within the binding pocket. By identifying which atom gives rise to each signal and how its intensity changes upon binding, we can infer the binding strength at that site, ultimately obtaining an epitope map of the ligand.

To achieve automatic assignment, we developed a matching algorithm formulated within a Bayesian framework. A similar approach has been successfully applied in the context of solid-state NMR [70], demonstrating the robustness of this methodology for complex spectral assignments. The algorithm matches two sets of chemical shifts: the experimental shifts measured from the photo-CIDNP spectra and the predicted shifts derived from the ligand structure. The predicted chemical shifts are annotated with the label of the hydrogen atom responsible for each signal. Once a correspondence between predicted and experimental shifts is established, the label of the predicted shift is transferred to the matched experimental one, thus completing the assignment.

The prediction of chemical shifts is performed by implementing a Graph Neural Network [128, 71], which takes as input a sparsified graph representation of the bidimensional structure of the ligand and outputs, for each hydrogen atom, the expected chemical shift for the ^{1}H NMR spectrum. The details of the network implementation will be presented in the next section. In the remainder of the present section, we will describe the Bayesian assignment algorithm under the assumption that the predicted chemical shifts and their corresponding hydrogen labels are already available.

The goal is to construct a bidirectional partial assignment between the hydrogen atoms of the ligand and the experimentally observed hyperpolarised signals. Let H denote the set of non-equivalent hydrogen atoms in the ligand molecule, with cardinality $N = |H|$, and let S denote the set of experimentally measured chemical shifts, with cardinality $M = |S|$. The objective is to establish a correspondence between a subset of H and a subset of S.

The assignment is *partial* in the sense that not all hydrogen atoms or experimental signals are necessarily involved, and *bidirectional* in the sense that each selected hydrogen atom is associated with at most one experimental chemical shift, and each experimental shift is associated with at most one hydrogen atom.

This formulation is motivated by two main considerations. First, not all hydrogen atoms in the ligand give rise to detectable hyperpolarised signals: depending on the experimental conditions and on their position within the binding pocket, some nuclei may not exhibit measurable signal enhancement and should therefore remain unassigned. Second, some experimentally observed chemical shifts may originate from artefacts or from spurious hyperpolarised signals introduced during sample preparation or data acquisition. Such signals do not correspond to any ligand atom and should therefore be left unassigned.

By explicitly allowing for unassigned atoms and unassigned experimental signals, the Bayesian formulation avoids enforcing incorrect correspondences and provides a physically meaningful mapping between the ligand structure and the experimental observations, which is essential for a reliable interpretation of signal quenching and for the estimation of binding strength.

From the set of chemical shift predictions, associated uncertainties, and hydrogen labels $\{y_i, \sigma_i, l_i\}$ provided by the graph neural network, we model the probability of observing an experimental chemical shift y arising from hydrogen atom l_i as a Gaussian distribution centered at the predicted shift $\hat{y}_i$ and with standard deviation σ_i, which reflects the prediction uncertainty:

$$p_i(y) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y-\hat{y}_i)^2}{2\sigma_i^2}\right]. \tag{7.13}$$

To account for experimentally observed signals that do not correspond to any hydrogen atom in the ligand, we introduce an explicit null-assignment. The probability of observing a chemical shift that is not associated with any predicted hydrogen is defined as a constant penalty,

$$p_{\text{none}} = \exp(-\lambda), \tag{7.14}$$

where $\lambda = 9/2$ is chosen to correspond to a 3σ threshold of the Gaussian model. With this choice, assigning a signal to the null hypothesis carries the same likelihood cost as assigning it to a predicted shift lying three standard deviations away from the observation.

On the basis of the above definitions, a score matrix is constructed to represent the likelihood associated with each potential pairing between predicted and experimental chemical shifts. Rows and columns whose entries do not exceed the null-assignment probability p_{none} are discarded, as they indicate predicted nuclei or experimental signals for which assignment is less probable than non-assignment.

Given the vector of experimentally observed chemical shifts $\boldsymbol{y} = \{y_j\}$, the probability that an observed shift y_j originates from hydrogen atom i, or from no hydrogen atom ($i = \text{none}$), is obtained by normalizing the corresponding likelihoods:

$$p(y_j \mid i) = \frac{p_i(y_j)}{\sum_k p_k(y_j) + p_{\text{none}}}. \tag{7.15}$$

An assignment $\boldsymbol{a}$ is defined as a vector specifying the correspondence between hydrogen nuclei in the molecule and the experimentally observed chemical shifts. For each

hydrogen nucleus i, the entry $a_i = j$ indicates that nucleus i is assigned to the experimental shift y_j, while $a_i = \text{none}$ denotes that nucleus i is not associated with any observed signal.

Given an assignment $\boldsymbol{a}$, the likelihood of observing the full vector of experimental chemical shifts $\boldsymbol{y}$ is expressed as the product of the individual assignment probabilities,

$$p(\boldsymbol{y} \mid \boldsymbol{a}) = \prod_i p(y_{a_i} \mid i), \tag{7.16}$$

where $p(y_{a_i} \mid i)$ is defined by the probabilistic model introduced above.

Application of Bayes' theorem yields the posterior probability of an assignment a given the observed vector of chemical shifts $\boldsymbol{y}$:

$$p(\mathbf{a}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\mathbf{a})p(\mathbf{a})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\mathbf{a})p(\mathbf{a})}{\sum_{\mathbf{a}'} p(\boldsymbol{y}|\mathbf{a}')p(\mathbf{a}')} \tag{7.17}$$

where $p(\mathbf{a})$ is a prior distribution over assignments, which we assume to be uniform for all valid assignments. An assignment is considered valid if it satisfies the bidirectional partial assignment constraints outlined above; otherwise, its prior probability is set to zero. To guide the assignment it is possible to include additional physical-chemical knowledge into the prior distribution. In the present case we know that with photo-CIDNP NMR hydrogens attached to heteroatoms (e.g., OH, NH, SH) do not show any hyperpolarization. Therefore, we set the prior probability of assigning any experimental shift to such hydrogens to zero.

The posterior distribution $p(\mathbf{a} \mid \boldsymbol{y})$ is defined over the space of assignments between n labels, corresponding to individual nuclei or groups of chemically equivalent nuclei, and m experimentally observed spectral signals. No assumption is made on the relative values of n and m. Under the bidirectional partial assignment constraints introduced above, each label can be assigned to at most one signal, while up to m signals may remain unassigned. The total number of valid assignments consistent with these constraints is given by

$$\sum_{k=0}^{\min(n,m)} \binom{n}{k}\binom{m}{k} k!, \tag{7.18}$$

where k denotes the number of assigned label–signal pairs, $\binom{n}{k}$ selects which k labels are assigned, $\binom{m}{k}$ selects which k signals receive an assignment, and $k!$ counts the number of bijections between the two chosen sets. The remaining $n-k$ labels and $m-k$ signals are left unassigned. This quantity increases combinatorially with n and m, making exhaustive enumeration of all assignments rapidly intractable for realistic spectra. For typical photo-CIDNP fragment screening spectra, n and m are on the order of 5 to 20, yielding a number of valid assignments on the order of 10^4 to 10^{10}.

To control this combinatorial growth while preserving the Bayesian formulation, global assignment generation is preceded by a pruning step. Only atom–shift pairings with sufficiently large posterior probability are retained. Specifically, for each nucleus, a candidate shift is kept only if its assignment probability exceeds a fraction $1/p_{\text{cutoff}}$ of the most probable shift for that nucleus. Since only predicted and experimental shifts that are spectrally close to one another have non-negligible assignment probability, this pruning naturally divides the global assignment problem across the full spectrum into a set of local subproblems, each involving only a small group of predicted and experimental shifts that are in close spectral proximity. In practice, each such local subproblem contains well

below ten signals, typically between two and five, making exhaustive enumeration within each subset computationally feasible.

The cutoff factor p_{cutoff} is treated as a tunable parameter and was found empirically to perform best when set to 1000 for the large majority of cases considered in this work.

Eq. (7.17) assigns a unique probability to each possible assignment of the measured chemical shift vector $\mathbf{y}$ to the set of atomic environments. This probability provides the appropriate probabilistic metric for comparing different global assignments. A more concise and interpretable quantity is the marginal probability that atom i is associated with shift j. This quantity is obtained by summing over all assignments $\mathbf{a}$ that contain the specific pairing $a_i = j$:

$$p(a_i = j \mid \boldsymbol{y}) = \frac{\sum_{\mathbf{a}} \delta_{a_i,j} p(\mathbf{a} \mid \boldsymbol{y})}{\sum_{\mathbf{a}} p(\mathbf{a} \mid \boldsymbol{y})}. \tag{7.19}$$

7.1.5 Chemical shift prediction with Graph Neural Network

The aim is to predict proton chemical shifts in order to obtain a set of reference values against which the experimentally observed signals can be matched, thereby enabling automated signal assignment. The Larmor frequency of a nucleus in NMR spectroscopy is determined by its surrounding electronic environment. Within a molecule, the local magnetic field experienced by each nucleus differs from the externally applied field: in some regions it is enhanced, while in others it is reduced due to the circulation of electrons. The precession frequency of each nucleus is therefore proportional to the locally modified magnetic field generated by the valence and bonding electrons. As a result, the Larmor frequency varies from one atomic site to another, depending on both the position of the nucleus within the molecular structure and the details of the electronic distribution. Therefore, any reliable prediction of chemical shifts must be grounded in an accurate representation of the molecular structure, since the arrangement of atoms and the nature of their bonding directly determine the electronic environment surrounding each nucleus.

Molecular structural information can be expressed in a natural and intuitive way through a graph-based representation, where atoms correspond to nodes and bonds to edges. This approach effectively preserves the intrinsic connectivity and topology of the molecule, providing a flexible and informative framework for describing its internal relationships. Graph neural networks have already demonstrated strong performance in accurately predicting a wide range of molecular and chemical properties. [129]

We represent each molecule as an undirected graph $G = (V, E)$, where V and E denote the set of nodes and the set of edges, respectively. The molecular structure is embedded into a sparse representation where only heavy atoms (see Tab.7.1) are represented as nodes in the network. Hydrogen atoms are treated implicitly and incorporated into the node features of their adjacent heavy atom. This simplification significantly reduces the size of the graph, lowering computational costs and memory requirements, while still capturing the necessary information, as hydrogen atoms bonded to the same heavy atom are symmetric within the molecular topology. On the other hand, edges are included only between nodes that correspond to atoms directly connected by a chemical bond.

Each node and edge in the molecular graph is associated with a corresponding feature vector (see Tab.7.1) that encodes relevant chemical information. The node feature vector $v_i \in V$ describes the properties of the i-th heavy atom, encompassing its atom type, formal charge, degree, hybridisation state, implicit valence, and the number of attached

hydrogen atoms. It also includes stereochemical and electronic characteristics, such as chirality, electron-donating or -accepting behaviour, aromaticity, and ring membership, as well as the number of rings to which the atom belongs for each ring size. Similarly, the edge feature vector $e_{i,j} \in E$ encodes the properties of the chemical bond connecting the i-th and j-th heavy atoms. These features capture the bond type, stereochemistry, conjugation, and whether the bond is part of a ring structure.

Table 7.1: Node and Edge Features for Sparse Molecular Graph Representation

Category	**Feature Type**	**Dim**	**Values**
Node Features	Atom Type	33	H, Li, B, C, N, O, F, Na, Mg, Al, Si, P, S, Cl, K, Ti, Zn, Ge, As, Se, Br, Pd, Ag, Sn, Sb, Te, I, Hg, Tl, Pb, Bi, Ga, Pt
	Formal Charge	6	1, 2, 3, -1, -2, -3
	Atom Degree	8	1, 2, 3, 4, 5, 6, 7, 8
	Hybridization	5	SP, SP2, SP3, SP3D, SP3D2
	Total Hydrogens	4	1, 2, 3, 4
	Total Valence	8	1, 2, 3, 4, 5, 6, 7, 8
	Donor/Acceptor	2	donor, acceptor
	Chirality	2	CW, CCW
	Ring Sizes	6	3, 4, 5, 6, 7, 8
	Aromaticity/Ring	2	is_aromatic, is_in_ring
	Total	**76**	
Edge Features	Bond Type	5	single, double, triple, aromatic, dative
	Stereochemistry	2	cis, trans
	Ring/Conjugation	2	in_ring, is_conjugated
	Total	**9**	

The nodes corresponding to heavy atoms bonded to hydrogen nuclei are annotated with the chemical shifts of the adjacent protons; these nodes constitute only a subset of the total set of nodes in the molecular graph. The label set can be defined as $Y = \{y_i \mid v_i \in V_a\}$ where $V_a \subset V$ represents the subset of nodes for which chemical shift annotations are available.

We implemented a message-passing neural network based on a path-augmented graph transformer architecture [71, 130, 131], designed to operate on molecular graphs and to perform node-level regression by learning context-aware atomic representations. The message passing function proceeds iteratively, progressively refining each node's embedding by aggregating and updating information obtained from neighboring nodes in the graph. Each input node vector $\mathbf{v}_i \in V$ is mapped into M distinct node representations through a set of embedding functions ϕ^m, yielding the embeddings $\mathbf{h}_i^{(0),1}, \ldots, \mathbf{h}_i^{(0),M}$, as follows:

$$\mathbf{h}_i^{(0),m} = \phi^m(\mathbf{v}_i), \quad m = 1, \ldots, M, \tag{7.20}$$

where each ϕ^m is parametrized as a single linear layer followed by a ReLU activation.

At the l-th message-passing step, the node embeddings are updated via a self-attention mechanism that assigns learned attention coefficients to neighboring nodes. This mechanism allows the model to weight contributions from different neighbors and to capture complementary relational patterns within the embedding space. The unnormalized attention scores are computed as

$$s_{i,j}^{(l),m} = \psi^{(l)}\left(\mathbf{h}_i^{(l-1),m}, \mathbf{h}_j^{(l-1),m}, \mathbf{e}_{i,j}\right), \quad \text{for } (i,j) \in E, \tag{7.21}$$

$$\alpha_{i,j}^{(l),m} = \frac{\exp\left(s_{i,j}^{(l),m}\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(s_{i,k}^{(l),m}\right)}, \tag{7.22}$$

where $\psi^{(l)}$ is a feedforward neural network parametrized as a sequence of linear transformations applied to the source node, destination node, and edge features, followed by a ReLU activation and a final linear projection to a scalar.

At the l-th message-passing step, the node embedding is obtained by combining a transformed version of the previous-layer embedding with an attention-weighted aggregation of neighboring node and edge features, together with a residual contribution from the initial embedding. The update rule is given by

$$\mathbf{h}_i^{(l),m} = \mathrm{ReLU}\left(\mathbf{h}_i^{(0),m} + \mathbf{W}_n^{(l)}\mathbf{h}_i^{(l-1),m} + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}^{(l),m}\left(\mathbf{M}_d^{(l)}\mathbf{h}_j^{(l-1),m} + \mathbf{M}_e^{(l)}\mathbf{e}_{i,j}\right)\right), \tag{7.23}$$

where $\alpha_{i,j}^{(l),m}$ denotes the attention coefficient associated with edge (i,j), and $\mathbf{W}_n^{(l)}$, $\mathbf{M}_d^{(l)}$, and $\mathbf{M}_e^{(l)}$ are trainable linear transformations. The explicit addition of the initial embedding $\mathbf{h}_i^{(0),m}$ introduces a residual connection that propagates low-level node information to all subsequent layers, stabilizing training and enabling deeper message passing while mitigating over-smoothing.

After L successive message-passing steps, the node embeddings have been iteratively refined to incorporate information from increasingly larger graph neighborhoods. The final node representation is then obtained by aggregating the M head-specific embeddings through averaging, yielding

$$\mathbf{h}_i^{(L)} = \frac{1}{M}\sum_{m=1}^{M} \mathbf{h}_i^{(L),m}. \tag{7.24}$$

The predicted chemical shift associated with the i-th node is obtained via a node-level readout function r_n, implemented as a feedforward neural network that maps the learned node representation to a scalar output. The node-level readout incorporates also a graph-level embedding $\mathbf{g}$ that produces a compact, permutation-invariant representation of the molecular graph, so that the model can complement local atomic information with global structural context. The node- and graph-level readout operations are defined as

$$\hat{y}_i = r_n\left(\mathbf{h}_i^{(L)}, \mathbf{v}_i, \mathbf{g}\right), \tag{7.25}$$

$$\mathbf{g} = r_g\left(\left\{(\mathbf{h}_i^{(L)}, \mathbf{v}_i) \mid \mathbf{v}_i \in V\right\}\right), \tag{7.26}$$

where $\hat{y}_i$ denotes the predicted output for node i, $\mathbf{h}_i^{(L)}$ denotes the final hidden representation of node i after L message-passing layers, $\mathbf{v}_i$ represents the original node features, and $\mathbf{g}$ is the graph-level context embedding.

The training dataset $D = \{(G_t, Y_t)\}_{t=1}^{N}$ is constructed from the NMRShiftDB2 database [132, 133] by processing N molecular entries with the RDKit Python library [134]. For each

molecule, the corresponding molecular graph $G_t = (V_t, E_t)$ is generated by extracting node- and edge-level features in the required graph representation and pairing them with the associated chemical shift labels Y_t. To reduce the influence of potentially incorrect or inconsistent chemical shift assignments present in the database, we aggregated all available reference chemical shifts corresponding to each hydrogen nucleus and used their median value as the target label in the analysis.

The loss function which is optimized during training is defined as

$$\mathcal{J} = \frac{1}{\sum_{t=1}^{N} |V_t^a|} \sum_{t=1}^{N} \mathcal{L}\Big(Y_t, \hat{Y}_t\Big), \tag{7.27}$$

$$\mathcal{L}\Big(Y_t, \hat{Y}_t\Big) = \sum_{i\,:\,\mathbf{v}_i \in V_t^a} |y_i - \hat{y}_i|, \tag{7.28}$$

where $V_t^a \subseteq V_t$ denotes the subset of NMR-active nodes corresponding to hydrogen nuclei in graph G_t, y_i and $\hat{y}_i$ are the reference and predicted chemical shifts associated with node i, respectively, and $\mathcal{L}$ computes the total node-wise absolute deviation over the ground truth values.

7.1.6 Binding strength estimates

At this stage of the automatic processing algorithm, the hyperpolarised signals have been detected and assigned to the specific ligand atoms responsible for their generation. This assignment is first performed on the ligand-only spectrum and subsequently transferred to the spectrum acquired in the presence of the protein, using the multiplet ranges predicted by MuSe Net. Establishing a consistent atom–signal correspondence across the two experimental conditions enables a direct, signal-wise comparison of spectral amplitudes.

Differences in signal amplitude reflect changes in the local magnetic environment experienced by the ligand nuclei, which may arise upon interaction with the protein. In particular, ligand binding can alter the efficiency of radical pair formation and spin dynamics underlying the hyperpolarization mechanism, thereby reducing the observed signal intensities. These amplitude variations form the basis for identifying binding-related effects and for selecting ligand atoms that are most strongly involved in the interaction.

To quantify this effect, we compute the binding contrast [135] that characterizes the amplitude differences between the two spectra:

$$C_{\text{bind}} = 1 - \frac{\Gamma_{L+P}}{\Gamma_L}, \tag{7.29}$$

where Γ denotes the integral of a specific signal in the NMR spectrum, reflecting the overall intensity of that resonance. The ratio $\frac{\Gamma_{L+P}}{\Gamma_L}$ expresses the relative change in the integrated signal intensity between the ligand in the presence of the protein, Γ_{L+P}, and the ligand-only spectrum, Γ_L. By construction, $C_{\text{bind}} = 0$ indicates no binding, as the signal intensity is unaffected by the presence of the protein, while $C_{\text{bind}} = 1$ corresponds to complete quenching of the hyperpolarised signal, indicating strong binding. In this formulation, C_{bind} serves as a signal-wise descriptor that quantifies the extent to which the spectral amplitude is quenched upon ligand–protein complex formation, providing insight into how strongly each atom of the ligand interacts with the surrounding environment within the protein's binding pocket.

For effective application of photo-CIDNP in fragment screening, it is essential to maximize the binding contrast between ligand signals. Introducing a relaxation delay in the experimental pulse sequence enhances this contrast by exploiting differences in relaxation behavior between bound and unbound ligand states.

Within this framework, it is possible to define a quantitative measure, referred to as the ligand score [13], which is proportional to the contribution of the protein to the overall relaxation process. This quantity reflects how the presence of the protein modulates the relaxation behavior of the ligand. In particular, the transverse relaxation rate R_2 of the small molecule in the presence of the protein can be expressed as

$$R_{2,+P} = p_{free} R_{2,free} + p_{bound} R_{2,bound}, \tag{7.30}$$

where p_{free} and p_{bound} denote the fractional populations of the ligand in the free and bound states, respectively, defined as the fraction of ligand molecules occupying each state at equilibrium, and $R_{2,\text{free}}$ and $R_{2,\text{bound}}$ are the corresponding transverse relaxation rates in each state. Under the assumption of a large excess of ligand relative to the protein, such that $p_{\text{free}} \approx 1$, the effective transverse relaxation rate in the presence of protein, denoted by $R_{2,+P}$, can be approximated as

$$R_{2,+P} \approx R_{2,free} + p_{bound} R_{2,bound} = R_{2,free} + R'_{2,bound}. \tag{7.31}$$

The effect of T_2 relaxation on the observed peak intensity can, in general, be expressed by

$$I(t) = I_0 e^{-R_2 t}, \tag{7.32}$$

where I_0 denotes the peak intensity measured in the absence of a relaxation delay, and $I(t)$ represents the peak intensity measured after a delay time t. The corresponding normalized peak intensity as a function of time can thus be expressed as

$$\frac{I(t)}{I_0} = e^{-R_2 t}. \tag{7.33}$$

The ratio of normalized peak intensities acquired at two distinct delay times, t_1 and t_2 (e.g. 10 ms and 200 ms), can be related through the following expression:

$$\frac{I_{t_2}}{I_{t_1}} = \frac{e^{-R_2 t_2}}{e^{-R_2 t_1}} = e^{-R_2 \Delta t}. \tag{7.34}$$

Finally, the ligand score Q is defined as

$$Q = \frac{I_{t_2,+P}}{I_{t_1,+P}} \frac{I_{t_1,free}}{I_{t_2,free}} = \frac{e^{-R_{2,+P}\Delta t}}{e^{-R_{2,free}\Delta t}} = \frac{e^{-(R_{2,free}+R'_{2,bound})\Delta t}}{e^{-R_{2,free}\Delta t}} = e^{-R'_{2,bound}\Delta t}. \tag{7.35}$$

To have a measure that is 0 for no ligand-protein interaction and 1 for perfect binding, we can consider the Q_{bind}:

$$Q_{\text{bind}} = 1 - Q = 1 - \frac{\Gamma_{L+P}}{\Gamma^*_{L+P}} \frac{\Gamma^*_L}{\Gamma_L}, \tag{7.36}$$

where the asterisk denotes the integral of the corresponding signal measured in the reference spectrum. In practice, the reference condition is chosen as the experiment acquired without any relaxation delay.

An additional advantage of Q_{bind} over C_{bind} is that the internal normalization inherent in its definition reduces sensitivity to small concentration mismatches between the ligand-only and ligand–protein measurements. This normalization also mitigates the influence of noise, making Q_{bind} generally more robust to experimental variability and signal fluctuations.

To classify a ligand–protein pair as a successful binding event (hit), both C_{bind} and Q_{bind} are subsequently binarized using an empirically defined threshold:

$$\begin{aligned} C_{\text{bind}},\, Q_{\text{bind}} \geq 0.2 &\rightarrow \text{Hit}, \\ C_{\text{bind}},\, Q_{\text{bind}} < 0.2 &\rightarrow \text{No Hit}. \end{aligned}$$

Although increasing the relaxation delay enhances binding-induced contrast, it simultaneously reduces the overall signal-to-noise ratio $SN = \frac{I}{\sigma_{\text{noise}}}$ as a result of relaxation-driven signal decay. Determining the optimal experimental protocol for fragment screening therefore necessitates quantifying the signal-to-noise ratio and selecting the experimental condition that jointly optimizes contrast and signal-to-noise performance. To this end, the proposed analysis pipeline performs this optimization automatically and suggests to the user the experimental protocol that best balances these competing effects.

7.2 Automatic fragment screening

7.2.1 Dataset for Statistical Evaluation

An accurate quantification of signal quenching, which constitutes the experimental signature of ligand–protein binding, depends critically on the reliable identification and characterization of hyperpolarised signals in the photo-CIDNP spectrum. Consequently, the primary objective in developing an algorithm for the automatic annotation of NMR signals and the subsequent classification of small molecules as binders or non-binders was the establishment of an optimized and reproducible protocol for hyperpolarised signal detection. This protocol was systematically evaluated on a dataset comprising paired conventional NMR and photo-CIDNP spectra of 96 small molecules from the NexMR AG library, thereby enabling a systematic assessment of the intrinsic characteristics of the dataset, including signal-to-noise variability, line-shape distortions, and inter-sample spectral heterogeneity. In the following, we will refer to this dataset as the CIDNP dataset. For each sample, the following experiments were recorded: a standard ^{1}H NMR experiment (dark spectrum), and a single-scan ^{1}H NMR experiment acquired after 2 s of continuous irradiation [11] (light spectrum). The spectrometer ran at $600 MHz$ and the photo-CIDNP experiments were performed using continuous light irradiation provided by a 1.5 W laser operating at a wavelength of 520 nm. All photo–CIDNP measurements in the CIDNP dataset were performed under standardized buffer conditions. The samples were prepared in 20 mM sodium phosphate buffer containing 50 mM NaCl at pH 7.4. In all experiments, 10% (v/v) D_2O was added to provide a field–frequency lock signal. Fluorescein was employed as the photosensitizer at a concentration of 10 μM. For oxygen removal, the samples further contained 3 mM deuterated glucose,together with 200 nM glucose oxidase and 200 nM catalase. This enzymatic oxygen scavenging system ensures efficient depletion of dissolved oxygen, which is required to preserve the efficiency and reproducibility of the photo–CIDNP effect.

A distinctive feature of this dataset concerns the preparation of the ligand stock solutions. The compounds were dissolved at 100 mM concentration in non–deuterated DMSO.

Upon dilution into the NMR sample, this resulted in a comparatively intense residual DMSO resonance. Together with the resonances originating from the buffer components, the residual DMSO gives rise to characteristic signals at well–defined chemical shifts in the NMR spectra. Although these signals are not influenced by the photo–CIDNP effect, *i.e.* their amplitudes do not increase in the light spectrum relative to the corresponding dark spectrum, they represent systematic spectral contributions that must be explicitly handled by the algorithm.

In practice, these solvent and buffer resonances can be substantially more intense than the hyperpolarised signals of interest. As a consequence, they may partially mask or distort weak CIDNP–enhanced resonances located on their shoulders, thereby complicating reliable detection and quantitative assessment of small hyperpolarised peaks.

Following the characterization of the CIDNP dataset and the validation of the hyperpolarised signal detection protocol, the statistical evaluation of the automated ligand–protein binding detection workflow was carried out on an independent screening dataset. This second dataset comprised small molecules tested against an undisclosed biomolecular target provided by NexMR AG. In contrast to the CIDNP dataset, which was primarily used to establish and calibrate the signal detection procedure, the screening dataset served to assess the performance of the complete pipeline under realistic fragment screening conditions. The dataset consists of ^{1}H NMR measurements acquired under multiple experimental conditions for each fragment, both in the absence and in the presence of the target protein. Besides the dark and light spectra acquired in the same way as for the CIDNP dataset, a CIDNP-PEARLScreen pulse sequence was acquired with $200ms$, and $800ms$ relaxation delays [13]. This makes up for a total of 8 NMR traces for each fragment–target pair.

For the screening dataset, samples were prepared in 25 mM Tris buffer at pH 7.5, containing 300 mM NaCl, 1 mM TCEP, 0.1% glycerol, and 10% D_2O, and comprised 30 μM of the compound under investigation together with 15 μM fluorescein, 100 nM glucose oxidase, 100 nM catalase, and 2 mM deuterated glucose.

The reduced ligand concentration has a substantial impact on the NMR signal, primarily through a decrease in the signal–to–noise ratio. This reduction renders the analysis considerably more challenging, as all stages of the pipeline, from peak detection to intensity quantification, must operate under conditions of diminished spectral contrast. In particular, the reliable identification of weak hyperpolarised signals and the precise estimation of binding–induced quenching become increasingly sensitive to noise fluctuations. Moreover, the intense resonances originating from the buffer components induced pronounced baseline distortions, which substantially complicated global phasing of the spectra. As a consequence, a dedicated pre–processing strategy was required to achieve reliable phase correction and to recover symmetric, pure absorption peak lineshapes, thereby preventing systematic distortions from propagating to subsequent peak detection and quantification steps.

7.2.2 Identification of an Optimal Preprocessing Strategy

As a preliminary validation study, the proposed phase-correction algorithm was applied to the CIDNP dataset. A visual inspection of the corrected spectra indicates that phase distortions were consistently and substantially attenuated across the dataset. In particular, improvements were observed even in challenging scenarios in which the molecular resonances of interest exhibited much lower amplitudes than the matrix signals, as

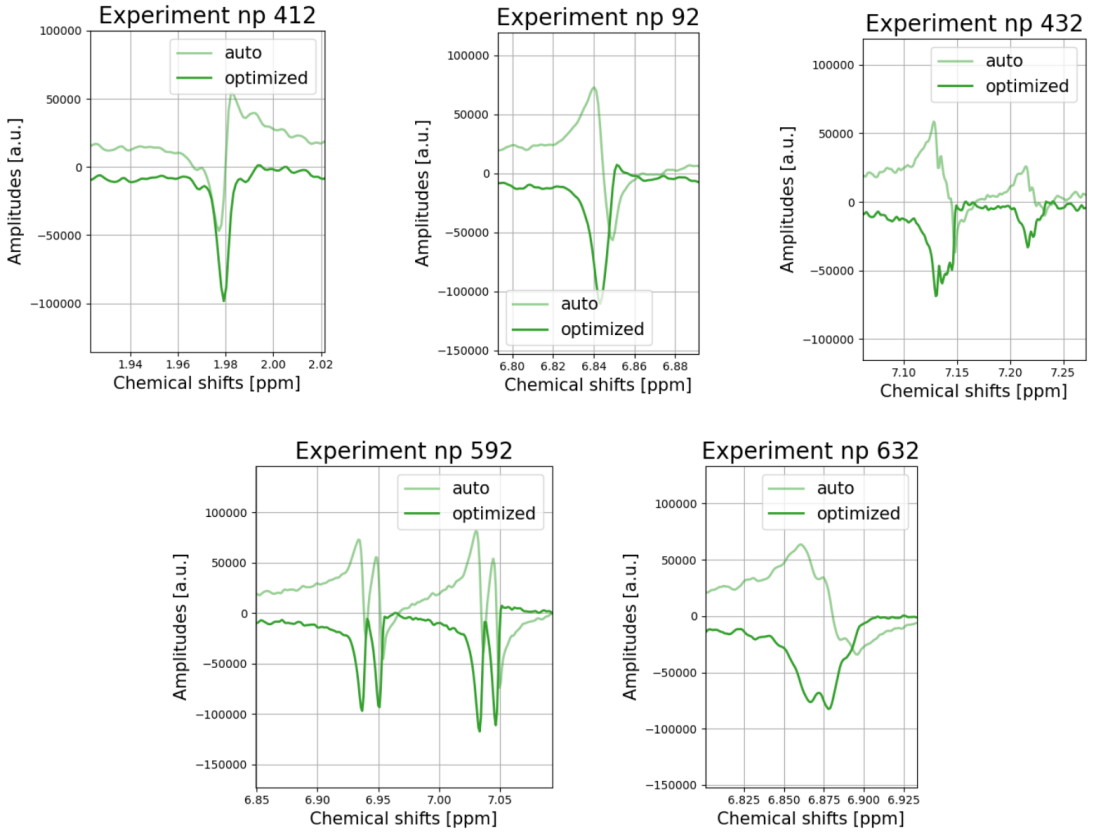


Figure 7.3: Comparison between the ACME algorithm for phase correction and the optimised phase correction procedure proposed in this work for photo-CIDNP data. Shown are five representative signal regions illustrating the improvement obtained with the proposed approach, in particular with respect to enhanced lineshape symmetry and the reduction of residual dispersion contributions compared to the ACME-corrected spectra.

well as in cases involving negative (emissive) peaks. In these situations, conventional phase-correction procedures often yield residual dispersive components or asymmetric line shapes, whereas the proposed method produced spectra with improved absorptive character.

In Fig. 7.3 a representative comparison between the standard ACME phase-correction algorithm [76] and the optimised procedure proposed in this work is shown for several signals from the CIDNP dataset.

Despite the implementation of the correction algorithm, the screening dataset exhibited persistent residual distortions, most likely related to strong baseline distortions introduced by the buffer signals and the reduced signal–to–noise ratio inherent to the screening conditions. Although a formal causal analysis was not performed, these remaining dispersive components introduce systematic errors during hyperpolarised signal selection by shifting the apparent peak maxima and biasing positional detection. More critically, residual phase errors can generate artificial negative lobes; because photo–CIDNP spectra intrinsically contain emissive signals, the automated routine may misclassify such artefacts as genuine hyperpolarised peaks.

A single–point calibration strategy, in which one representative spectrum was manually phased and the resulting parameters were applied uniformly to the full library, proved insufficient and led to inconsistent results across samples. In practice, reliable

screening requires strict phase coherence between ligand–only and ligand–protein spectra, since even small phase differences propagate into systematic variations in peak shape and amplitude, thereby compromising intensity–based binding metrics.

To reduce these effects, magnitude–mode processing was adopted, thereby removing the phase degree of freedom and yielding strictly positive line shapes for all spectra. Although this transformation modifies the analytical line shape, leading to broader peaks, and changes the noise statistics from Gaussian to Rician, which reduces to Rayleigh noise in the absence of signal, it provides a more stable and reproducible basis for automated peak detection.

Empirically, the procedure that yielded the most symmetric peaks consisted of first phasing the complex spectrum using manually optimized phase parameters. This step was required in order to perform an effective baseline correction on both the real and imaginary components prior to magnitude computation (see Figure 7.4). At first sight, this may appear counterintuitive, since magnitude processing was introduced precisely to avoid phase sensitivity. However, phase correction is necessary to enable proper baseline correction; if baseline distortions are not removed beforehand, they propagate through the nonlinear magnitude transformation and manifest as asymmetric peak shapes. This asymmetry has been previously noted by De B. Harrington *et al.* [69], who anyhow advocated the use of magnitude spectra for pattern recognition purposes. In particular, it is essential to apply baseline correction independently to both the real and imaginary components of the complex spectrum. This represents a departure from standard processing workflows, in which the imaginary part is typically discarded after phasing, and therefore not processed at all.

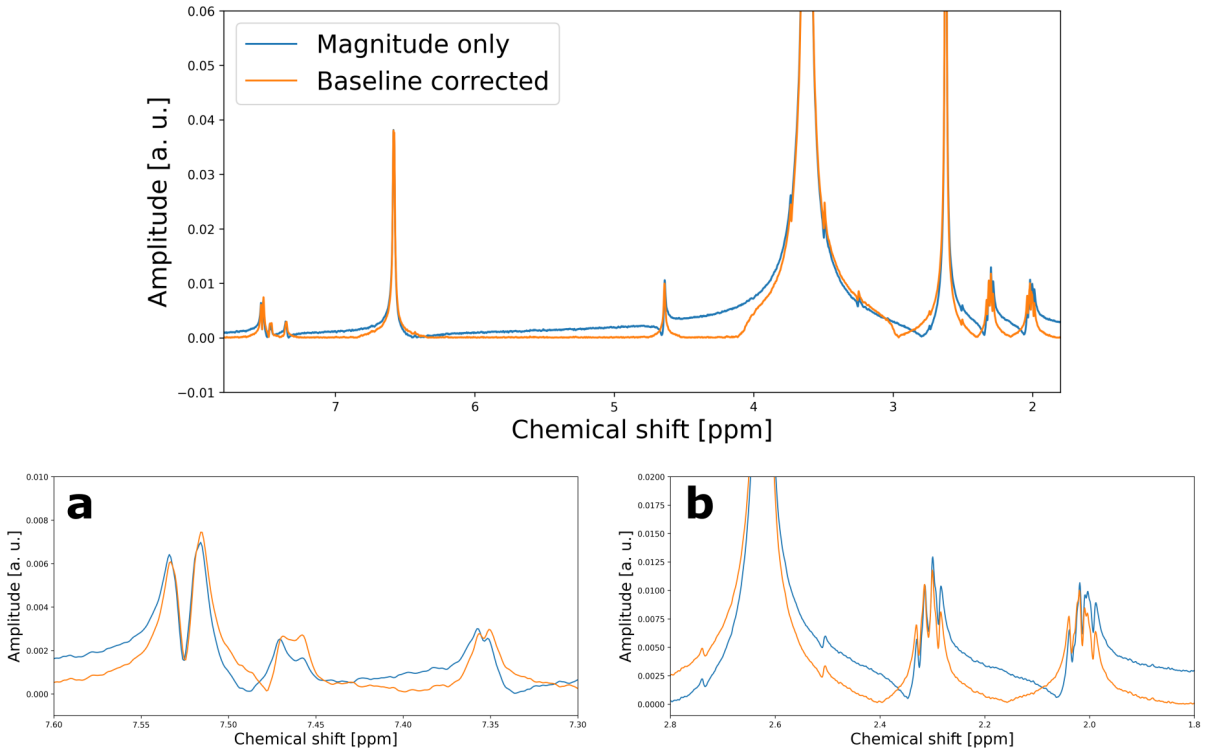


Figure 7.4: Comparison between two magnitude-mode representations of the same ^{1}H NMR spectrum from the screening dataset. The blue trace corresponds to the magnitude spectrum obtained directly from the Fourier-transformed data, while the orange trace shows the result obtained after applying baseline correction prior to the magnitude computation. The comparison illustrates how baseline correction reduces distortions in the magnitude spectrum and results in peaks with a more symmetric lineshape.

7.2.3 Hyperpolarised signal detection

For each molecular fragment included in the present study, photo-CIDNP active resonances had previously been examined by experienced NMR spectroscopists from the partner company NexMR AG. These hyperpolarised peaks were systematically identified in the light spectra through visual inspection and comparison with the corresponding dark reference spectra, taking into account intensity enhancements, emissive behaviour, and possible peak displacements. Where feasible, the resonances were assigned to specific nuclei on the basis of the known molecular structure. The resulting annotations, including peak positions and assignment metadata, were curated and stored in the NexMR internal database. In the present work, these expert-produced annotations were used as a standard reference to assess the performance of the proposed automatic detection algorithm.

As described in the previous section, the automatic identification of hyperpolarised signals proceeds in two sequential stages. First, individual peaks are detected by a peak-picking procedure based on local maxima criteria, line-shape consistency, and a heuristic comparison of signal intensities between the dark and light spectra. In a second step, the detected peaks are grouped into multiplets by means of MuSe Net, which performs a data-driven clustering of resonances according to learned spectral patterns.

After the initial peak-picking stage, a non-negligible number of spurious detections is observed. These false detections predominantly correspond to small-amplitude noise

fluctuations that satisfy the local maxima criterion but do not represent genuine resonances. The subsequent MuSe Net–based clustering step exhibits increased robustness with respect to such noise-induced artefacts. In practice, this second stage acts not only as a multiplet reconstruction module but also as a filtering mechanism that removes a substantial fraction of the spurious peak predictions generated in the first stage. This effect was verified quantitatively (see Table 7.2).

For each fragment, the set of automatically selected peaks was compared with the corresponding expert reference assignments. On this basis, individual detections were classified as true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). The true negative peaks in this context are all the peaks that have been picked but not selected as hyperpolarised through comparison of dark and light spectra. From these counts, dataset-level detection statistics were computed, including balanced accuracy, precision, recall, and the F_1 score thereby providing a quantitative assessment of the algorithm's performance relative to expert annotations.

Before applying MuSe Net, the algorithm tends to substantially overselect hyperpolarised peaks. Although this strategy ensures that virtually all CIDNP-active signals are captured, it also leads to a considerable number of false-positive detections. Visual inspection shows that many of these spurious selections correspond to noise artefacts. In several cases, they are located on the shoulders or tails of intense resonances, where small fluctuations can appear artificially enhanced relative to the dark spectrum.

The detection accuracy remains high(see Table 7.2). This value is predominantly driven by the large number of true negatives, which exceed the other confusion matrix entries by approximately one order of magnitude. The high accuracy value reflects the strong performance of the automatic procedure in correctly matching peaks between dark and light spectra and in reliably identifying the small subset of hyperpolarised signals within a vast majority of unaffected resonances.

After introducing MuSe Net clustering, the selection becomes more stringent and better aligned with expert annotation (see Figure 7.5).

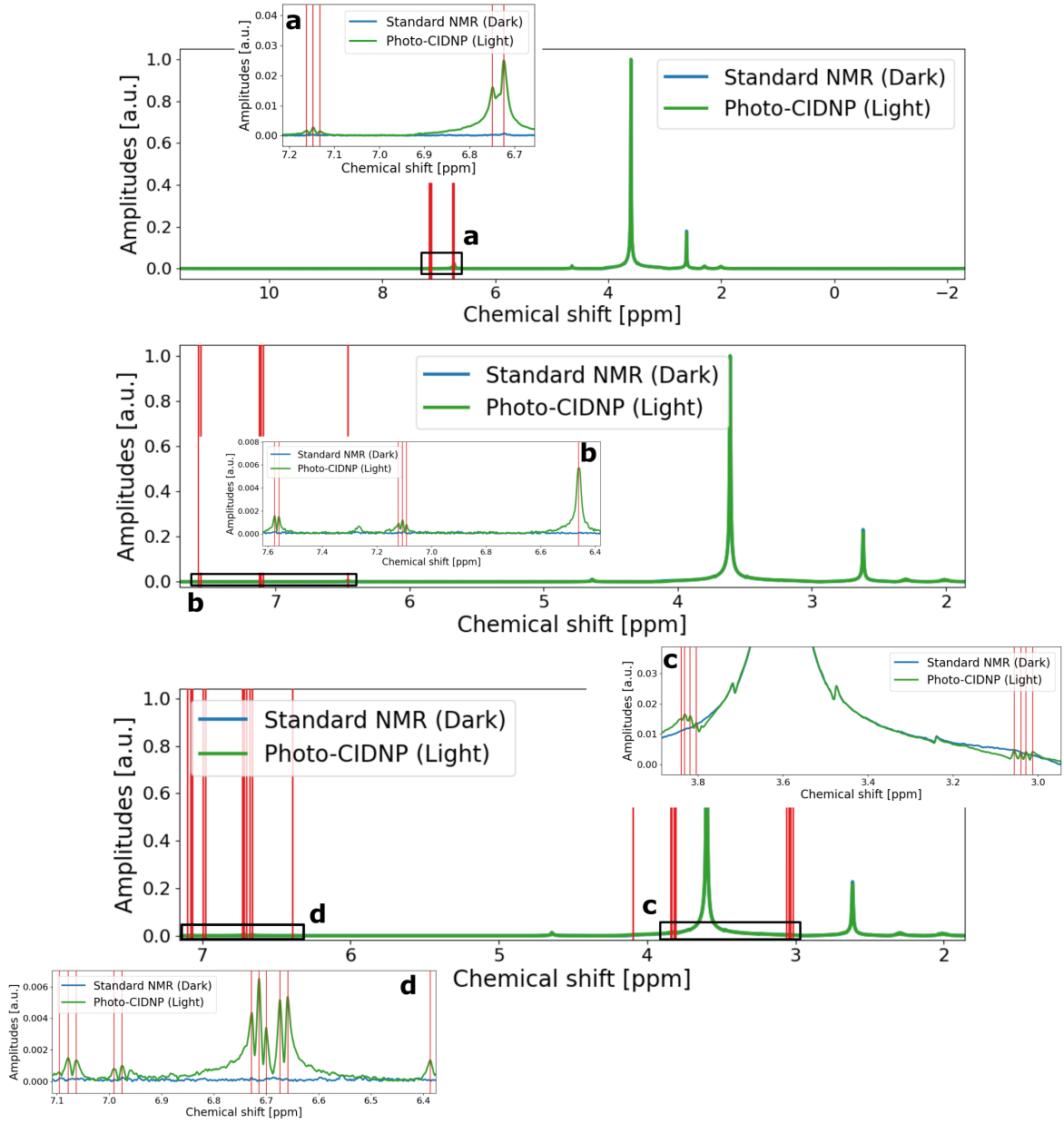


Figure 7.5: Illustration of the automated peak picking procedure applied to three photo-CIDNP ^{1}H NMR spectra from the screening dataset. These examples illustrate the ability of the procedure to discriminate hyperpolarized signals from the non-hyperpolarized signals in the spectra. In particular, Insets b and d demonstrate robust detection under reduced signal-to-noise conditions, whereas Inset c shows successful identification of a hyperpolarized signal located on the shoulder of a strong buffer peak.

The clustering step effectively suppresses these artefactual detections while retaining the majority of genuinely hyperpolarised signals. This is quantified by the increase of 32.5% in precision and of 19.5% in F1 score. The filtering operated by MuSe Net inevitably leads to a decrease in recall; however, the overall balance between sensitivity and selectivity is substantially improved, supporting the robustness of the method for automated photo-CIDNP screening applications.

To further assess the robustness of the automatic algorithm, we evaluated its performance under more challenging screening conditions. In the screening dataset, the

Table 7.2: Comparison of the performance of hyperpolarised signal detection on the CIDNP dataset, evaluated before and after MuSe Net clustering. Reported metrics quantify the impact of the clustering stage on sensitivity to CIDNP-active peaks and on the control of false-positive detections.

Dataset	Accuracy	Precision	Recall	F1-score
Before MuSe Net	**0.979**	0.413	**1.000**	0.584
After MuSe Net	0.908	**0.739**	0.824	**0.779**

ligand concentration is three times lower than in the CIDNP dataset, leading to a reduced signal-to-noise ratio. In addition, the signal-to-noise ratio decreases further with increasing transverse relaxation delay. This setting therefore provides a stringent test of the algorithm in conditions closer to realistic fragment screening scenarios.

We specifically investigated whether using the magnitude spectrum instead of the real part could improve detection performance. The improvement is clearly quantifiable across all evaluated metrics and for all tested relaxation delays (see Table 7.3). For $T_2 = 0\,\text{ms}$, $200\,\text{ms}$, and $800\,\text{ms}$, the magnitude spectrum consistently outperforms the real spectrum in terms of accuracy, recall, precision, and F1-score. In particular, the gain in precision is substantial, indicating a marked reduction in false-positive detections, while recall remains comparatively high even at longer relaxation delays.

A key factor contributing to this improvement is the more symmetric lineshape in the magnitude spectrum, which reduces artefactual detections arising from suboptimal phasing. In the real spectrum, imperfect phase correction can distort peak shapes and artificially enhance shoulders or tails, thereby promoting spurious peak selections. The magnitude representation mitigates these effects, leading to a more stable and reliable identification of genuinely hyperpolarised signals.

Notably, for the case without relaxation delay, the recall obtained using the magnitude spectrum is even higher than that achieved on the CIDNP dataset, despite the lower ligand concentration. At the same time, the precision remains comparable, demonstrating that the method maintains strong discriminative power even under reduced signal-to-noise conditions. Overall, these results confirm the advantage of using the magnitude spectrum over the real part for automated hyperpolarised signal detection.

7.2.4 Mapping binding effects onto the ligand structure

Signal assignment was performed automatically using a Bayesian matching algorithm (sec. 7.1.4) that links experimentally observed resonances to the corresponding nuclei in the ligand structure. The experimental inputs are the chemical shifts extracted through the hyperpolarised signal selection procedure described above, while the reference values are the chemical shifts predicted from the molecular structure.

The prediction performance of the graph neural network (sec. 7.1.5) for chemical shift estimation was assessed on the NMRShiftDB2 [132, 133] dataset, which contains approximately 54 000 entries, using a ten-fold cross-validation protocol. The dataset was randomly partitioned into ten mutually exclusive subsets of approximately equal size. At each iteration, one subset was held out as the test set, while the remaining nine subsets were used for training. This process was repeated such that each subset served once as the test set, and the resulting performance metrics were averaged over all folds to obtain a robust estimate of predictive accuracy.

Table 7.3: Comparison of the performance of hyperpolarised signal detection on the screening dataset between real and magnitude spectral representations, evaluated at three transverse relaxation delays ($T_2 = 0$, 200, and 800, ms). The metrics (accuracy, recall, precision, and F1-score) quantify the relative sensitivity to CIDNP-active peaks and the robustness against false-positive detections as a function of spectral representation and relaxation delay.

Method	Accuracy	Recall	Precision	F1
$T_2 = 0\,\text{ms}$				
Real spectrum	0.842	0.726	0.242	0.363
Magnitude spectrum	**0.904**	**0.874**	**0.673**	**0.761**
$T_2 = 200\,\text{ms}$				
Real spectrum	0.804	0.649	0.240	0.350
Magnitude spectrum	**0.842**	**0.741**	**0.682**	**0.710**
$T_2 = 800\,\text{ms}$				
Real spectrum	0.754	0.538	0.262	0.352
Magnitude spectrum	**0.831**	**0.711**	**0.678**	**0.694**

Predictive performance on the test data was quantified using two standard error measures, namely the mean absolute error (MAE) and the root mean square error (RMSE). Both metrics were computed on a per-nucleus basis over the set of NMR-active (hydrogen) nuclei $|V_t^a|$ of molecule t in the test set:

$$\text{MAE} = \frac{1}{\sum_{t=1}^{N}|V_t^a|}\sum_{t=1}^{N}\sum_{i\,:\,\mathbf{v}_{t,i}\in V_t^a}\left|y_{t,i}-\hat{y}_{t,i}\right|, \tag{7.37}$$

$$\text{RMSE} = \sqrt{\frac{1}{\sum_{t=1}^{N}|V_t^a|}\sum_{t=1}^{N}\sum_{i\,:\,\mathbf{v}_{t,i}\in V_t^a}\left(y_{t,i}-\hat{y}_{t,i}\right)^2}, \tag{7.38}$$

where y_i and $\hat{y}_i$ are respectively the label and the predicted chemical shift for nucleus i.

Beyond cross-validation on NMRShiftDB2, the trained graph neural network was additionally evaluated on the fragment-screening dataset of photo-CIDNP ^{1}H spectra. The SD files containing the two-dimensional molecular structures of the ligands were processed using RDKit [134] in order to extract a sparse molecular graph representation. In this representation, atoms correspond to nodes and covalent bonds to edges, with associated atom- and bond-level features derived from the structural information encoded in the SD files. In Fig. 7.6, several representative ligands from the fragment-screening dataset are shown together with their corresponding graph representations.

Table 7.4: Predictive performance of different models for ^{1}H chemical shift estimation. Errors are reported in ppm. HOSE: Hierarchical Organisation of Spherical Environments; GCN: Graph Convolutional Network; FCG: Fully Connected Graph network; SGNN: Scalable Graph Neural Network. The performance values for HOSE and GCN were taken from [136], those for FCG and SGNN (NMRShiftDB2) from [71]. A dash indicates that the corresponding metric was not reported.

Model	MAE	RMSE
HOSE	0.33	–
GCN	0.28	–
FCG	0.22	–
SGNN (NMRShiftDB2)	0.22	0.49
SGNN (ours - NexMR AG)	**0.18**	**0.34**

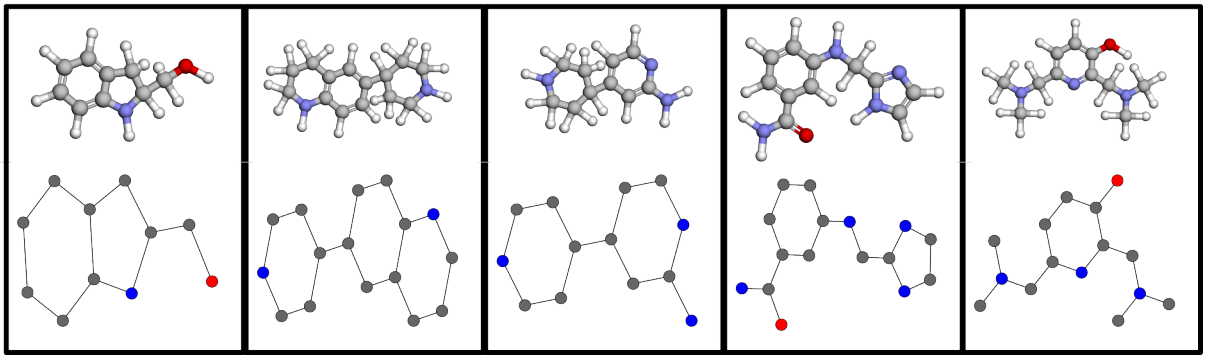


Figure 7.6: Examples of ligand molecules from the screening dataset together with their corresponding molecular graph representations used as input to the graph neural network.

This external evaluation assesses the model's ability to generalise to molecular fragments not present in the training dataset. Because the graph neural network operates on atom- and bond-level descriptors of the molecular graph, predictions are based on local chemical environments rather than on memorisation of specific structures. Performance on previously unseen fragments therefore provides an empirical test of the transferability of the learned structure–shift relationships. When applying the graph neural network trained on NMRShiftDB2 to the fragment-screening dataset without further fine-tuning, the predictive accuracy remained high and, in fact, surpassed that of previously reported algorithms evaluated under comparable conditions (see Table 7.4). The improvement over earlier approaches suggests that the combination of graph-based message passing and data-driven parameter optimisation provides a more expressive model of local chemical environments than fragment-based or shallower graph methods.

Accurate chemical shift prediction with limited error is a critical prerequisite for reliable signal assignment. Since the Bayesian matching procedure relies directly on the agreement between experimental and predicted shifts, prediction inaccuracies propagate into the assignment likelihood and can increase ambiguity or lead to incorrect mappings.

For each ligand, the algorithm produces two output files (see Fig. 7.7). The first contains the set of all tested global assignments together with their associated posterior probabilities. The second reports the marginal probabilities for individual nucleus–signal correspondences, obtained by summing over all compatible global configurations. This dual output enables both a ranking of complete assignment hypotheses and a local as-

```
Unambiguous assignments:
H16/H17          3.01
    H19          6.66
    H21          6.39

Ambiguity between H13/H14, H15
Corresponding shifts: 3.83
Mapping 2 distributions on 1 shifts (2) possible assignments
H13/H14          None          3.83
    H15          3.83          None
      P  50.00179606%  49.99820394%

Ambiguity between H18, H20
Corresponding shifts: 7.08, 7.12
Mapping 2 distributions on 2 shifts (2) possible assignments
    H18          7.12          7.08
    H20          7.08          7.12
      P  72.16857127%  27.83142873%
```

Label	3.01	3.83	6.39	6.66	7.08	7.12	None
H13/H14	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%	50.00%
H15	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%	50.00%
H16/H17	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
H18	0.00%	0.00%	0.00%	0.00%	27.83%	72.17%	0.00%
H19	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
H20	0.00%	0.00%	0.00%	0.00%	72.17%	27.83%	0.00%
H21	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%

Figure 7.7: (a) All possible assignments with their posterior probability and (b) marginal probabilities for nucleus-signal correspondences.

sessment of confidence at the level of individual nuclei.

The assignments, together with their respective probabilities, are reported in the output file and are organised according to the corresponding local assignments. These local assignments serve as the building blocks for constructing the space of possible global assignments. By combining assignments defined on subgroups of contiguous signals, the algorithm assembles consistent global solutions. This subdivision into subproblems reduces the combinatorial complexity of the overall matching procedure. In favourable cases, the assignment is unambiguous, namely when a single experimental shift is compatible with exactly one predicted shift within the admissible tolerance. This is the case of hydrogens $H16/H17$, $H19$, and $H21$ in the example (see Fig. 7.7). In other instances, an experimental shift may be compatible with multiple predicted shifts (and conversely, a predicted shift may correspond to several experimental candidates), as in the case of hydrogen $H13/H14$ and $H15$ which are competing for the same experimental shift. In such ambiguous situations, the algorithm evaluates all consistent combinations and reports the most probable global matchings together with their associated probabilities.

To quantitatively assess the performance of the automatic assignment algorithm, a subset of the full fragment-screening dataset was selected for evaluation. Specifically, we restricted the analysis to 50 molecules for which a complete and univocal expert assignment was available, ensuring a well-defined reference for comparison.

Assignment performance was evaluated using a ranking-based criterion derived from the marginal posterior probabilities. For each experimental signal, the algorithm provides a ranked list of candidate nucleus assignments together with their marginal probabilities. We computed the proportion of cases in which the expert reference assignment was contained within the n most probable marginal candidates, with $n \in \{1, 2, 3\}$.

Formally, let $\mathcal{A}_i^{(n)}$ denote the set of the n highest-ranked candidate assignments for experimental signal i, ordered by decreasing marginal probability, and let a_i^{ref} denote the corresponding reference assignment. The top-n accuracy is then defined as

$$\text{Acc}_n = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}\Big(a_i^{\text{ref}} \in \mathcal{A}_i^{(n)}\Big), \tag{7.39}$$

where M denotes the total number of evaluated signals and $\mathbf{1}(\cdot)$ is the characteristic function.

Top-1 accuracy reflects fully correct marginal identification, while top-2 and top-3

accuracies quantify whether the correct assignment is retained among the most probable alternatives in ambiguous cases.

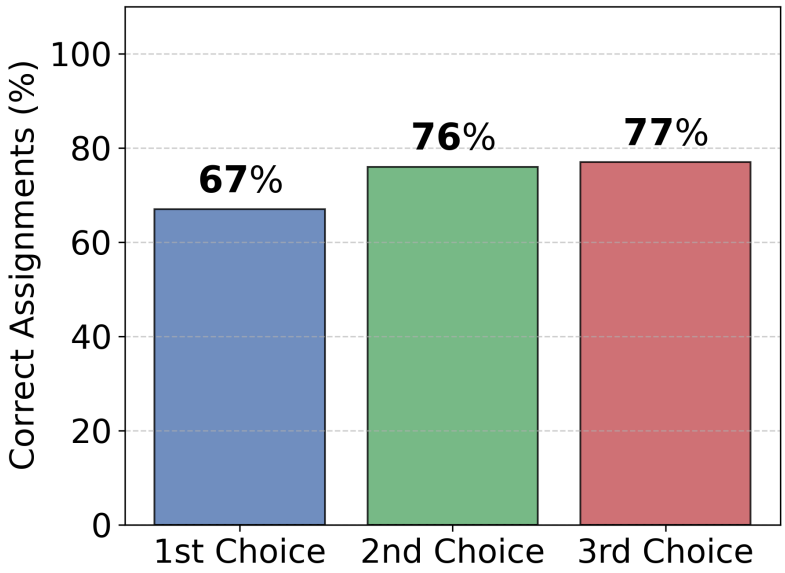


Figure 7.8: The top-n accuracy performance of the automatic assignment algorithm, with $n = 1, 2\,3$.

The ranking of assignments according to their posterior probability (see Figure 7.8) shows that the correct solution is identified as the most probable in 67% of cases. When extending the evaluation to the top two ranked assignments, the success rate increases to 76%, and reaches 77% when considering the top three. This indicates that the correct assignment is typically found among the highest-probability candidate solutions, supporting the reliability of the probabilistic scoring framework.

It should be noted that the current assignment framework exploits only the distance between predicted and experimentally observed chemical shifts. Performance could be further improved by incorporating additional information about the resonances, such as their multiplicity, which would provide complementary constraints and reduce ambiguity in cases where multiple candidate assignments yield similar shift distances.

From an application perspective, the atom assignment step plays a central role in translating the spectral observations into structural insight. Once each experimental signal has been assigned to a specific nucleus, the corresponding C_{bind} and Q_{bind} values can be mapped directly onto the ligand structure. Rather than characterizing binding as a single global quantity, this yields a spatially resolved picture of the binding profile, highlighting which parts of the molecule are most strongly affected upon interaction with the protein and which remain largely unperturbed.

7.2.5 Binding candidates selection

The method developed in this work quantifies ligand–protein interaction through the two scalar observables introduced in the previous sections, C_{bind} and Q_{bind}. These quantities are derived by comparing the photo–CIDNP enhancements observed in spectra of the ligand alone with those recorded in the presence of the target protein.

C_{bind} and Q_{bind} are computed for each signal location detected by MuSe-Net. The signal quenching (i.e., the reduction in signal intensity upon binding) is estimated either considering the individual peak intensities or the multiplet integral. These two strategies constitute alternative estimators of the same physical effect: the peak–height approach emphasizes local amplitude variations, whereas the integral–based approach captures the cumulative enhancement distributed across the entire multiplet envelope, with all peaks within a multiplet sharing the same integral–estimated value. The algorithm outputs a file reporting C_{bind} and Q_{bind} for each hyperpolarized peak. In the case of intensity-based measures, individual peak contributions can be aggregated into per-multiplet measures either by averaging over the multiplet or by selecting the peak with the highest signal-to-noise ratio. Finally, we select the signal displaying the largest signal quenching: its C_{bind} and Q_{bind} are then used for classifying the molecule as a successful binder.

The observable C_{bind} is evaluated for all relaxation delays T_2. In contrast, when measuring Q_{bind}, the signal ratio is computed using the zero–delay experiment as reference, and therefore the Q_{bind} is only estimated for experiments with $T_2 > 0$ delays.

The quantities C_{bind} and Q_{bind} are defined operationally and should be interpreted as empirical screening scores that reflect relative irradiation–induced enhancements. They are not direct estimates of thermodynamic binding constants. Their numerical values depend on experimental parameters such as irradiation time, ligand and protein concentration, and relaxation delay. Although multiple relaxation delays are considered, comparisons are performed only within experiments acquired under identical delay conditions, thereby ensuring internal consistency across samples.

Binding assessment is implemented by thresholding the scalar scores. Two independent classifications can be obtained: one based on C_{bind} and one based on Q_{bind}. Since both observables aim to capture the same physical phenomenon, consistency between the two classifications is expected under ideal conditions. In practice, discrepancies are not uncommon, even under manual annotation. For example, C_{bind} may be reduced in experiments without transverse relaxation delay $T_2 = 0$, whereas Q_{bind} measured at longer delays can suffer from reduced signal–to–noise ratios, which limits quantitative accuracy. For this reason, the two criteria are evaluated separately rather than combined into a single joint decision rule.

For C_{bind}, a molecule is classified as a Hit if its binding contrast exceeds a threshold θ_C, and as No Hit otherwise. Independently, for Q_{bind}, a molecule is classified as a Hit if its binding quenching score exceeds a threshold θ_Q, and as No Hit otherwise.

In addition to these outcome classes, a third category is defined to capture spectra in which no hyperpolarised signals are detected. Such cases may arise from experimental issues (e.g. sample preparation errors or insufficient ligand concentration) or from signal levels below the detection threshold. These samples are assigned to a no CIDNP class prior to the evaluation of binding scores. Introducing this category prevents confusion of technical failure with true absence of binding and improves interpretability of the screening results.

Under this formulation, the pipeline functions as a three–class classifier, distinguishing between: (i) absence of detectable CIDNP signal (no CIDNP), (ii) measurable enhancement consistent with binding (Hit), and (iii) measurable but sub–threshold enhancement (No Hit). Running the complete workflow, including hyperpolarised peak detection, peak assignment, and calculation of the binding measure, on the screening dataset required 8.79 minutes.Ultimately, this automated framework acts as a highly efficient prioritization filter to enable downstream hit-validation. By making the first critical selection,

the algorithm yields a manageable list of candidate fragments that can subsequently undergo visual inspection to confirm the algorithmic hits. From there, the pipeline allows researchers to confidently subject these selected fragments to orthogonal validation. Testing them in secondary, independent assays definitively proves that the observed binding is a genuine molecular interaction, rather than a false positive driven by other phenomena such as fragment aggregation.

The automated fragment classification was benchmarked against the C_{bind} and Q_{bind} annotations provided manually by expert spectroscopists at NexMR AG. Only molecules with unambiguous annotations were included in the evaluation (see Table 7.5).

Table 7.5: Annotation statistics of the fragment screening dataset. Rows correspond to the binding criterion, and columns report the number of molecules classified as No CIDNP, HIT, and NO HIT, respectively, using a threshold of $\theta = 0.2$.

Criterion	**No CIDNP**	**HIT**	**NO HIT**
C_{bind}	60	37	179
Q_{bind}	60	43	173
C_{bind} & Q_{bind}	60	17	153

For each measure configuration, a decision threshold on the corresponding binding score was optimised. Classification performance was evaluated using a 3×3 confusion matrix reflecting the three classes introduced above (No CIDNP, Hit, No Hit). From this confusion matrix, the performance metrics described in Chapter 4 were computed, including precision, recall, and the F_1 score for the Hit class. The optimal measurement type and threshold were selected by maximising the F_1 score of the Hit class.

Both for C_{bind} and Q_{bind}, the automatic measuring procedure that showed the highest agreement with the manual annotations was based on peak intensity measurements, selecting the highest peak within each multiplet.

For C_{bind}, the optimal configuration employed the pulse sequence without T_2 delay and a decision threshold of 0.24. For Q_{bind}, the optimal configuration corresponded to the pulse sequence with the longest T_2 delay and a threshold of 0.2.

At the optimal threshold, the algorithm retrieves 86.1% of annotated hits using C_{bind} and 88.2% using Q_{bind} (see Table 7.6). The recall increases to 91.7% when the C_{bind} threshold is set to 0.2, the value used during manual annotation. The precision, however, is suboptimal, particularly for C_{bind} screening at threshold 0.2, as the algorithm tends to overselect potential binder candidates. Although an ideal classifier would achieve both high precision and high recall, fragment screening prioritises sensitivity over specificity: it is preferable to retain a larger set of candidates and ensure that no true binders are discarded, provided that the selected molecules undergo further validation. In the present dataset, the automatic procedure reduced the number of molecules requiring further validation by 82% for both Q_{bind} and C_{bind} at threshold 0.24, and by 75% for C_{bind} at threshold 0.2. Combined with the substantial reduction in annotation time, from several days of manual photo-CIDNP analysis to less than 10 minutes of computation, the approach represents a significant acceleration of the fragment screening pipeline.

The high recall and moderate precision suggest that the algorithm systematically overestimates binding scores relative to the manually retrieved values. This tendency can be further quantified by evaluating C_{bind} and Q_{bind} as continuous quantities rather than binary classifications, treating binding assessment as a regression problem. This formulation is conceptually more demanding: whereas classification only requires a correct

Table 7.6: Comparison of the C_{bind} and Q_{bind} models for fragment screening. Reported values include hit-detection metrics (precision, recall, and F1-score for the *Hit* class), global classification metrics (balanced accuracy and macro-averaged F1-score), and regression metrics (RMSE and MAE).

Model	θ	$\text{Recall}_{\text{Hit}}$	$\text{Precision}_{\text{Hit}}$	F1_{Hit}	Bal. Acc.	F1_{Macro}	RMSE	MAE
C_{bind}	0.24	0.861	0.596	0.705	0.811	0.804	0.167	0.084
C_{bind}	0.20	0.917	0.493	0.641	0.806	0.772	0.167	0.084
Q_{bind}	0.20	0.882	0.588	0.706	0.847	0.751	0.165	0.102

decision relative to a chosen threshold, regression penalises deviations across the entire dynamic range of the binding scores, so that both systematic bias and variance directly affect the reported mean absolute error (MAE) and root mean square error (RMSE). The observed errors are on the order of 10%, with a notable fraction of cases in which the manual annotation is zero while the algorithm predicts a positive value. The agreement between automatic and manual annotations is inherently limited by the sequential nature of the pipeline: as C_{bind} and Q_{bind} are the result of multiple chained processing steps, inaccuracies introduced at each stage accumulate and propagate to the final binding score estimates. Additionally, certain cases may be handled differently by the automatic procedure compared to manual annotation; in particular, molecules annotated as non-binders (i.e., with a reference value of zero) may not be treated equivalently by the two approaches, potentially contributing to the observed errors. A more detailed quantitative analysis of these error sources is left for future work.

Chapter 8

Discussion

The main objective of the present doctoral project was to advance the automation of one-dimensional NMR data interpretation. Standard one-dimensional ^{1}H NMR experiments are characterized by comparatively low operational complexity and short acquisition times. In contrast to multidimensional pulse sequences, whose acquisition may take several hours or even days depending on resolution and sensitivity requirements, a conventional one-dimensional spectrum can typically be recorded within minutes. This substantial difference in measurement time translates directly into low experimental costs and high throughput. Consequently, one-dimensional ^{1}H NMR remains the method of choice in numerous applications spanning organic, inorganic, and biomolecular chemistry.

However, the reduced experimental burden is accompanied by intrinsic informational constraints. Multidimensional experiments encode correlations between nuclei along additional frequency axes. In particular, two-dimensional spectra display cross-peaks at coordinates defined by two independent frequencies, thereby directly revealing couplings between specific pairs of spins and significantly simplifying structural analysis [47]. In a one-dimensional experiment, the same relational information is not explicitly separated along an additional dimension. Instead, it must be inferred indirectly through the joint analysis of chemical shifts and resonance multiplicities. The identification of coupling partners therefore relies on the correct interpretation of splitting patterns and their quantitative parameters.

The absence of an additional frequency dimension also limits the ability to resolve spectral overlap. In two-dimensional spectra, dispersion along orthogonal axes frequently separates signals that would coincide in a one-dimensional projection. By contrast, in one-dimensional spectra, overlapping resonances may obscure accurate determination of chemical shifts, multiplicity patterns, scalar coupling constants, and relative intensities.

Considering the challenges and constraints imposed by the intrinsic nature of one-dimensional NMR data, yet recognizing its considerable scientific value, the present work aimed to investigate how molecular information can be systematically and reliably extracted from one-dimensional NMR spectra in the absence of prior structural knowledge. In this context, we pursued the task of identifying which features of the spectral signal can be systematically exploited, which limitations arise from the intrinsic configuration of one-dimensional NMR data, and which computational procedures can support the formulation of automated analysis methods. The goal was to develop algorithms capable of mimicking, in a formalized and reproducible manner, the operation that an expert spectroscopist typically performs during the manual annotation of NMR spectra.

These scientific questions formed the basis of the research activities carried out in

collaboration with Bruker Switzerland AG. We recognized the central role of multiplet splitting pattern analysis in determining the spin system of the molecule under investigation and proposed to formalize this task as a one-dimensional segmentation problem along the spectral axis [52, 83]. Within this framework, we developed MuSe Net, a supervised probabilistic deep learning architecture designed to classify multiplets in ^{1}H NMR spectra according to their splitting patterns. The model produces predictions at the level of individual spectral points, assigning to each position a label corresponding to the most probable multiplet class together with an associated confidence score that reflects the uncertainty of the classification. The uncertainty score provided by the model was also intended to assist in the identification of overlapping and composite signals, where the observed spectral patterns deviate from the standard multiplet classes expected for isolated spin systems and the classification task becomes intrinsically less well defined.

To train the model, a strategy was introduced for generating realistic synthetic spectra in a computationally efficient manner, avoiding the need for full simulations based on explicit molecular structures. The procedure relies on a consistent definition of the multiplet phenotype associated with each class, enabling the controlled generation of representative spectral patterns suitable for supervised training.

Eventually, another aspect of automation was addressed, namely the integration and correlation of information obtained from spectra recorded under different experimental conditions, including variations in the experimental setup and pulse sequences. This problem was investigated in the context of photo-CIDNP NMR fragment-based drug discovery within a research collaboration with NexMR AG.

A dedicated computational pipeline was implemented to automatically detect hyperpolarised signals and evaluate the attenuation of signal intensity between ligand-only and ligand–protein spectra, with the aim of identifying potential binders. The estimated quenching values were mapped onto the molecular structure of the ligand through the implementation of a Bayesian assignment algorithm. To enforce a consistent absorptive character of the signals across the spectra under comparison, a specific processing protocol was therefore prescribed to obtain symmetric peak shapes in the magnitude spectrum, thereby improving the robustness of automated peak detection.

8.1 Multiplet Segmentation as a Gateway to Multiple Spectral Descriptors

Multiplet analysis and classification with respect to their splitting patterns play a central role in reconstructing the spin system of the molecule under investigation. Indeed, the identification of multiplet regions provides access to several key spectral descriptors, including chemical shifts, scalar coupling constants, and signal integrals. Once the spectral interval corresponding to a multiplet has been localized, the chemical shift can be estimated as the midpoint of the identified range. Information on spin–spin couplings, which reflect the connectivity structure of the spin system, can then be inferred from the multiplicity and the relative positions of the individual peaks within the multiplet.

In this perspective, multiplicity and splitting patterns can also provide useful constraints for downstream spectral processing steps. In particular, they can guide peak picking and peak deconvolution procedures by imposing an upper bound on the number of expected peaks within a multiplet region. This information can help prevent overpicking caused by noise artefacts that may pass conventional amplitude thresholds, as well

as overfitting during deconvolution procedures. Furthermore, the integral of the signal associated with the multiplet provides an estimate of the abundance of the corresponding proton.

These different quantities can, in principle, be inferred starting from a single machine learning task, namely the semantic segmentation of the one-dimensional spectrum, which simultaneously provides the localisation of multiplet regions and their classification according to multiplicity. This is the strategy adopted in the MuSe Net framework.

Most traditional approaches reported in the literature do not provide the localization of the multiplet region itself and instead assume that it is already known. Deconvolution-based methods [32, 33], pattern-recognition approaches [34], and techniques based on inverting the splitting tree of the multiplet [35, 36, 37, 38] are all designed to operate on a predefined multiplet window. In other words, they require that the spectral interval containing the multiplet has already been identified before the analysis can be applied.

An attempt to extend the analysis to the entire spectrum was proposed by Griffiths et al. [39], who introduced a method to identify multiplets directly from a list of accurately detected peaks. In this approach, multiplets are recognized by exploiting positional symmetries among peaks. However, the method relies on the availability of a reliable peak table, which must be carefully curated to remove residual solvent signals and other artefacts. Consequently, even though the procedure is applied at the level of the full spectrum, it still depends on a prior localization step through peak detection.

Additional information about the signal must generally be provided. In the context of pattern-recognition approaches, the initial theoretical model must be carefully selected so that it adequately reproduces the experimental signal, particularly with respect to the assumed line shape. In methods based on the inversion of the splitting tree, the requirements are even more restrictive: besides the localization of the multiplet region, the positions of the visible peaks must be supplied, together with an estimate of the number of spins coupled to the nucleus under consideration. This information is necessary in order to define an upper bound on the number of peaks that may constitute the multiplet, including components that may not be directly observable due to symmetry or overlap. Moreover, these methods typically require a preliminary normalization of peak intensities before the analysis can be performed. As a consequence, they are not able to account for distortions such as the roof effect, where the relative intensities of the multiplet components deviate from the ideal symmetric pattern expected under first-order conditions. Finally, all these methods rely on recursive optimization procedures. In deconvolution-based approaches, the spectrum is progressively simplified through iterative fitting of line components. Pattern-recognition methods operate by optimizing the agreement between the experimental signal and a predefined theoretical pattern. Similarly, techniques based on the inversion of the splitting tree or on peak clustering identify multiplet structures by iteratively searching for symmetry relations among detected peaks. Consequently, the procedure must be executed repeatedly for each multiplet region and subsequently applied to all multiplets present in the spectrum. This iterative optimization, due to the non-convex nature of the problem, may lead the algorithm to converge to suboptimal solutions.

In contrast, MuSe Net performs detection and classification of signals across the entire spectrum simultaneously, without requiring the user to predefine the spectral regions containing individual multiplets. The localization of these regions is instead produced as part of the model output. The method operates directly on spectra with an arbitrary number of data points and requires only a single forward pass to obtain a classification

of the splitting patterns included in the training set. For uncertainty quantification, the inference procedure is repeated ten times, corresponding to the evaluation of each network within the ensemble. The variability among the predictions produced by the ensemble members is then used to estimate the classification uncertainty and to identify spectral regions where the signal is likely to correspond to overlapping or composite multiplets. Furthermore, the framework is able to handle several deviations from ideal first-order multiplet behaviour. In particular, MuSe Net can correctly classify signals affected by the roof effect without requiring additional procedures aimed at enforcing symmetry in the amplitudes of the multiplet components. This capability is illustrated, for example, by the correct prediction of doublet of doublets in Figure 5.15, as well as by additional examples involving dt and ddd patterns (see top row of Figures 5.4 and 5.5). The method also shows robustness in the presence of large variations in signal intensity, moderate baseline distortions, and small phase imperfections. An example of its ability to operate over a large dynamic range arises in the application of MuSe Net within the photo-CIDNP analysis pipeline, where hyperpolarised signals can in some cases be significantly weaker than the signals originating from compounds present in the sample preparation matrix.

The multiplet classification task can also be formulated within an object-detection framework, where multiplets are identified as bounding regions along the spectral axis. In this setting, localization is treated as a regression task involving two quantities: the position of the box centre (anchor) and its width, which together define the spectral interval associated with the multiplet [137, 138]. In addition to the localization step, each detected box is assigned a multiplicity label through a classification module. Consequently, the object-detection formulation decomposes the problem into three explicit tasks, two regression tasks for localization and one classification task for multiplicity, whereas the segmentation approach encodes the same information implicitly through the assignment of class labels along the spectral axis. One limitation occasionally observed in the segmentation framework is the appearance of fragmented predictions within a multiplet region. This effect is generally avoided in the box-based formulation, where the multiplet is represented by a single contiguous detection window. In practice, however, fragmentation tends to occur primarily for challenging resonances (see bottom row of Figures 5.4 and 5.5), which are also characterized by higher predictive uncertainty. In such cases, mild post-processing through filtering or region-merging procedures is typically sufficient to recover a consistent multiplet localization.

8.2 Representing Multiplet Phenotypes for Supervised Learning

Deep learning models are often described as data-driven methods, as they acquire the ability to perform a given task by extracting relevant statistical regularities from the training data. This contrasts with rule-based or model-based approaches, in which the algorithmic behavior is explicitly determined through predefined analytical rules or physical models. In data-driven methods, the instructions required to perform the task are not explicitly encoded in analytical rules, but are instead implicitly embedded in the examples presented during training, from which the model learns the relevant feature representations and decision boundaries. Consequently, the composition of the training dataset plays a decisive role in determining whether a deep learning approach can be successfully deployed in practical applications.

In the case of MuSe Net, this requirement translates into the need to construct a training dataset that provides an adequate representation of the spectral features, namely the multiplet structures observed in ^{1}H NMR spectra. Since the model must infer the rules that distinguish different splitting patterns directly from the examples provided during training, the dataset must contain sufficiently diverse and representative instances of each multiplet class. Only under this condition can the network learn decision boundaries that are both expressive and internally consistent, avoiding contradictory cues that could arise from poorly defined or insufficiently represented spectral patterns. Careful design of the training data is therefore essential to ensure that the model captures the characteristic features of the multiplet phenotypes and generalizes reliably to experimental spectra.

Defining what constitutes an adequate representation of the relevant spectral features is not straightforward. A natural starting point would be to assemble a very large dataset containing thousands of annotated examples. In practice, however, the construction of sufficiently large labeled datasets of experimental spectra is extremely challenging. The volume of annotated data typically required to train deep neural networks would imply the acquisition and expert annotation of a very large number of spectra, a process that is both time-consuming and costly. In addition, manual annotation is inherently subject to variability between experts, which may introduce inconsistencies in the labels and ultimately affect the reliability of the learning process.

To overcome these limitations, an automated procedure was implemented to generate synthetic spectral segments containing multiplets belonging to the eleven classes selected as the most frequently encountered in ^{1}H NMR spectra. Rather than relying on a full physical simulation of the spectrum starting from an explicit molecular structure, the procedure generates the signals by randomly sampling the relevant multiplet parameters within predefined ranges. This strategy considerably reduces the computational cost while still producing representative spectral patterns. An automatic labeling procedure was applied to each generated sample, enabling the production of a large and consistently annotated training dataset suitable for supervised learning.

To ensure an information-rich training set, the relevant multiplet parameters were sampled across extended ranges of values, generating spectra that exhibit a broad variety of peak spacings, coupling constants, linewidths, and amplitudes. The sampling intervals were deliberately chosen to extend somewhat beyond the values most frequently encountered in experimental spectra. This choice reflects the well-known tendency of deep learning models to interpolate reliably within the domain represented in the training data while performing less robustly when extrapolation is required. By covering a sufficiently wide region of the parameter space, the training procedure therefore facilitates the construction of expressive and stable decision boundaries for the classification task.

Distinguishing among splitting patterns in ^{1}H NMR spectra can be particularly challenging because all resonances ultimately consist of combinations of individual peaks and may therefore exhibit similar local features. To improve the internal coherence of the dataset and reduce ambiguities between classes, a regularization procedure was introduced during data generation. This procedure includes the exclusion of overlapping multiplets and a relabeling step applied when particular combinations of J-couplings produce peak configurations that resemble those of a different multiplet class, for example when symmetry or near-degenerate couplings effectively hide some peaks in the pattern. These measures help ensure that each class corresponds to a well-defined spectral multiplet phenotype.

Several factors contribute to the observed class imbalance. First, the baseline class

naturally dominates the dataset, since even in experimental spectra the regions containing signals occupy only a small fraction of the total frequency axis, while most spectral points correspond to baseline. This predominance of baseline points proved beneficial for the learning process, enabling the network to reliably distinguish noise from genuine signal and to identify the boundaries of signal regions with high accuracy. In addition, multiplet classes containing a larger number of peaks and coupling constants tend to extend over wider spectral intervals and therefore contribute a greater number of labeled points when a point-wise representation is adopted. At the same time, these complex multiplets are more likely to undergo relabeling into simpler coupling classes when particular combinations of J-couplings cause some peaks to merge or become effectively indistinguishable. Furthermore, because of their larger spectral extent, they also have a higher probability of being discarded during dataset generation, either because they produce overlapping peaks or because the multiplet does not fit entirely within the spectral window of the generated segment.

Despite this imbalance, the model achieved consistently high performance across all multiplet classes in synthetic spectra (see Figures 4.7 and 5.2, Table 4.6 and 5.1), indicating that the moderate disparity in class frequencies did not significantly affect the overall classification accuracy. Nevertheless, the unequal distribution of labeled points among classes should be taken into account during dataset construction and model implementation. If necessary, the representation of underrepresented classes can be increased by generating additional patterns, while overly frequent classes may be moderated by selectively reducing their occurrence in the training set. In addition, constraints can be introduced during the sampling of coupling constants to ensure that the J values remain sufficiently separated, thereby preventing the collapse of peaks in the splitting tree, which could otherwise transform multiplets characterized by multiple J-couplings into patterns resembling simpler one-coupling multiplets.

The resulting framework showed robust performance on experimental spectra (see Figures 4.9 and 5.3) acquired across a variety of application domains, including metabolomics, drug discovery, and materials science contexts such as lithium-ion battery research. The model generalized well to spectra originating from molecules of different structural complexity and recorded under varying acquisition conditions, maintaining accurate predictions even at relatively low signal-to-noise ratios. In some cases, the introduction of a small amount of additional white noise slightly improved the predictions (see Fig. 5.10), suggesting that weak noise may help suppress spurious correlations introduced by signal suppression procedures or other processing artefacts.

A limitation was observed primarily for more complex multiplet classes such as dt and ddd (see bottom row of Figures 5.4 and 5.5), where the experimentally observed phenotype occasionally resembled that of simpler splitting patterns. In these cases, the expert annotation was informed by knowledge of the molecular structure, information that was not available to the network, which relied exclusively on the spectral features present in the signal. For the dddd class, a more extensive statistical assessment on real experimental spectra would be necessary to reliably evaluate the model performance.

It is also noteworthy that, when applied to photo-CIDNP spectra without retraining, MuSe Net continued to provide meaningful predictions despite the presence of domain shifts relative to the training data. In particular, emissive peaks were largely ignored by the model, while correct classifications were still obtained when magnitude-mode spectra were used as input. This behaviour is not trivial, since the magnitude transformation modifies the statistical distribution of the noise and alters the peak shape. The fact

that the model remains effective under these conditions indicates a substantial degree of robustness with respect to variations in spectral representation and acquisition conditions.

8.3 Probabilistic Deep Learning for Multiplet Segmentation

The architecture of MuSe Net was designed to capture both local spectral features and their correlations along the frequency axis. At a first stage, the convolutional layers perform a form of local pattern inspection of the spectrum, extracting features at different length scales through the application of convolutional kernels of varying sizes. This operation allows the network to detect elementary spectral motifs, such as peak shapes and local peak configurations, while maintaining robustness to small variations in position and intensity.

The features extracted by the convolutional blocks are subsequently processed by a Long Short-Term Memory (LSTM) layer, which models the sequential dependencies along the spectral axis. This component enables the network to capture correlations between neighboring spectral regions, which are essential for recognizing multiplet structures whose defining characteristics arise from the relative arrangement of peaks rather than from isolated local features.

Finally, fully connected layers map the extracted representations into a higher-level feature space in which the classes become more separable, facilitating the final classification step. The combination of convolutional, recurrent, and dense layers has proven particularly effective for the analysis of one-dimensional signals and sequential data, where it allows the model to first extract local features and subsequently integrate them across the sequence to capture global structural patterns [53, 54, 55, 56, 57].

In recent years, Transformer architectures [139] have significantly influenced the field of deep learning, enabling models capable of processing long sequences and ultimately paving the way for large language models. A defining feature of Transformers is their ability to process input sequences in parallel and to directly model correlations between elements that may be far apart within the sequence through self-attention mechanisms. This represents an advantage compared to CNN–LSTM architectures, where information propagation follows the sequential structure of the model. At the same time, convolutional and recurrent architectures possess strong inductive biases that are particularly suitable for signal-processing tasks. Convolutional neural networks naturally capture translation invariance and local spatial patterns, while LSTM layers model sequential dependencies along the input signal. Although recurrent networks may suffer from vanishing gradients when modeling long sequences, Transformer architectures come with different trade-offs. In particular, self-attention introduces quadratic time and memory complexity, $\mathcal{O}(N^2)$, with respect to the sequence length N, whereas CNN–LSTM models scale linearly, $\mathcal{O}(N)$. Moreover, Transformers typically require larger datasets to learn spatial and sequential relationships that CNNs and LSTMs encode through architectural inductive biases. As a consequence, both training and inference may become computationally demanding unless substantial GPU parallelization is available. These considerations become particularly relevant when uncertainty quantification is required [140]. Approaches such as Monte Carlo dropout or deep ensembles rely on multiple forward passes during inference. In this setting, the computational demands of Transformer architectures can become considerable, whereas CNN–LSTM models remain more practical in many applications [141].

MuSe Net provides a concrete example of this, as its deep ensemble implementation is able to produce multiplet segmentation over an entire spectrum together with an associated confidence score in under one minute.

The ability to quantify uncertainty is in fact a crucial aspect of deploying deep learning models in scientific applications. While modern deep neural networks often achieve high accuracy, their probabilistic outputs do not necessarily provide reliable estimates of confidence. In a multiclass classification setting, for example, the softmax probabilities frequently assign a very high score to the predicted class even when the prediction is incorrect. In other words, the model may appear highly confident in decisions that are in fact wrong. From a statistical perspective, this behavior reflects the well-known issue of poorly calibrated predictive probabilities in deep learning models. Informally, one may say that neural networks tend to behave as if they were unable to admit their ignorance.

Such overconfident errors represent a critical limitation in applications where reliable decision-making is required. This problem has been widely discussed in contexts such as autonomous driving and medical diagnosis, where incorrect but confident predictions may lead to severe consequences. Although the consequences in scientific data analysis are typically less life-threatening than in safety-critical applications, the same principle applies: predictions should ideally be accompanied by an estimate of their reliability. In quantitative scientific workflows, the ability of a classification algorithm to provide meaningful confidence estimates constitutes an essential methodological requirement.

Despite the increasing tendency within the scientific community to adopt models that operate largely as opaque predictive systems [58, 59] calls for careful epistemiological reflection, developing strategies for uncertainty quantification is not only a matter of methodological rigor, but also of practical utility. Beyond increasing the trustworthiness of the predictions, uncertainty estimation can provide insight into the internal behavior of the model. By examining which predictions are associated with high or low uncertainty, it becomes possible to identify which spectral features are well captured by the learned representation and which configurations remain difficult for the network.

In the case of MuSe Net, the analysis of uncertainty also provided useful information about the origin of prediction errors. For synthetic spectra (see Figure 5.6), the dominant contribution was found to arise from aleatoric uncertainty, which reflects the intrinsic stochastic variability in the relationship between inputs and outputs and is therefore irreducible. In experimental spectra, by contrast, the main source of uncertainty was epistemic, which is associated with limited knowledge about the optimal predictive model and may be reduced by providing additional training data.

Uncertainty estimation also proved valuable for identifying resonances exhibiting spectral configurations that deviate from those represented in the training data. These include multiplets with convoluted splitting patterns not included among the training classes, as well as overlapping resonances, both of which can be regarded as out-of-distribution (OOD) signals. Once identified through elevated uncertainty scores, these challenging cases can be subjected to additional analysis or referred to expert inspection, thereby allowing the algorithm to effectively signal situations in which its prediction should not be trusted.

The effectiveness of OOD detection depends on how strongly the distance between in-distribution (ID) and OOD samples in feature space is reflected in the uncertainty estimates. In practice, adjusting the uncertainty threshold can improve the discrimination between reliable and unreliable predictions. However, in rare cases similar uncertainty values were observed for both ID and OOD multiplets, which limits the effectiveness of

the detection. No clear regularities were identified that could attribute this behavior to specific experimental conditions, and further investigation would be required to better understand these occurrences.

8.4 From Multiplet Classification to Structure Verification

MuSe Net provides a spectral segmentation that enables the retrieval of key spectral parameters. The chemical shift can be estimated from the central position of the segmented multiplet region, the proton number from the integral of the region, and the J–coupling constants by combining the predicted splitting pattern with the positions of the resolved peaks. Peak positions can be extracted using `mldcon`. These parameters can then serve as input for spectral simulation algorithms that fit the experimental spectrum and verify the corresponding spin system. As a proof of concept, we demonstrated this procedure on a simple molecule with a single spin system A_2X_3, where the extracted parameters were used as priors in a Bayesian fitting framework [114]. The reconstruction of the experimental spectrum from these parameters was highly satisfactory (Fig. 5.16), illustrating the feasibility of the approach.

This example, however, also highlights a fundamental limitation. The parameters obtained from MuSe Net do not specify the topology of the spin–spin coupling network, that is, which nuclei are mutually coupled. A tentative inference may be attempted when similar coupling constants appear in different multiplets, suggesting a possible connection between the corresponding spins. In the simple example considered here the solution was trivial because only two signals were coupled. For more complex molecules, however, this inference becomes considerably more difficult and the number of possible candidate spin systems grows rapidly.

An interesting continuation of this work would therefore consist in bridging this gap by automatically proposing plausible spin-system topologies for the interpretation of a given ^{1}H NMR spectrum. Such a task could be addressed using deep learning methods trained on data that encode the relationship between spectral parameters and molecular structure. One possible strategy is to exploit open-source NMR databases that provide chemical shifts, multiplicities, and proton counts [132, 142], while sampling additional parameters such as linewidth, lineshape, noise level, phase, and baseline distortions in the same fashion as in the MuSe Net training procedure. Alternatively, spin systems can be simulated directly and subsequently distorted to produce realistic spectra.

An example of this latter strategy is the MolDeTr framework [138], which employs a transformer-based architecture adapted for one-dimensional NMR signals. In this approach a convolutional backbone first extracts local spectral features, while transformer layers model long-range dependencies across the spectrum, enabling the identification of related multiplet patterns and chemical shifts. Chemical prior information, such as typical ranges of coupling constants and chemical shifts, is incorporated to guide the prediction process. The model is trained using a set-based bipartite matching loss similar to that employed in Detection Transformers, which enforces a one-to-one correspondence between predicted and true spectral features and avoids the need for heuristic post-processing. The resulting model directly outputs a structured representation of the spectrum, including the location of multiplet regions together with estimates of chemical shifts, coupling constants, and integrals. The coupling topology is still inferred as a post-processing step by

matching similar coupling constants across multiplets.

Future work should extend this framework to more complex molecules containing multiple interacting spin systems and should incorporate a more explicit quantification of predictive uncertainty. Finally, it is worth noting that analogous interpretation procedures, based on the characterization of spectral features through their position, integrals, and patterns, are widely employed in other spectroscopic techniques. Examples include mass spectrometry, where true peptide fragment peaks must be distinguished from chemical noise, Raman and infrared spectroscopy, where characteristic fingerprint bands are identified, and X-ray photoelectron spectroscopy, where peak positions and areas are used to determine oxidation states and elemental compositions. In this context, deep learning architectures similar to MuSe Net and MolDeTr, which excel at extracting molecular information from complex spectral data—possess significant potential for adaptation across these diverse spectroscopic domains.

8.5 Correlating Information Across Spectra for Binding Detection

An additional challenge arises when information must be extracted from multiple spectra acquired under different experimental conditions and the corresponding spectral features must be correlated to detect meaningful changes. This situation occurs, for example, in fragment-based drug discovery experiments monitored by NMR spectroscopy [143]. In this context, the interaction between a small molecule ligand and a target protein can be inferred by comparing spectra recorded in the absence and in the presence of the protein. Successful binding typically manifests as a reduction in the intensity of ligand resonances when the protein is added.

The computational pipeline developed in this work was specifically designed for photo-CIDNP NMR spectra, a technique that has recently been proposed as an effective method for detecting ligand–protein interactions in aqueous solutions at low concentrations and at room temperature [11, 12]. Photo-CIDNP exploits a light-induced hyperpolarization effect that strongly enhances the NMR signal of specific nuclei. As a consequence, a single scan is often sufficient to achieve a signal-to-noise ratio adequate for detection. This characteristic makes photo-CIDNP particularly attractive for high-throughput applications, as it allows rapid acquisition of spectra across large ligand libraries.

In typical fragment-screening experiments, a single protein target is tested against several hundred ligands in order to identify a limited number of potential binding candidates. Under these conditions, the availability of fast and reliable analysis procedures becomes essential in order to match the high experimental throughput enabled by the technique.

Several open-source and commercial platforms are available to assist researchers in the annotation and analysis of spectra acquired for fragment-based screening using commonly employed NMR techniques, such as ^{1}H, ^{19}F, STD, WaterLOGSY, $T_{1\rho}$, and T_2 relaxation experiments. In widely used commercial environments such as TopSpin [144], the analysis workflow remains largely centered on an interactive graphical user interface. In this setting, hit identification typically relies on manual inspection of the spectra, followed by annotation with a binary status indicating binding or no binding. Other platforms provide a higher degree of automation. For example, MNova [145] offers tools to manage the analysis workflow from raw spectral data to automated hit suggestions. Neverthe-

less, final validation still relies on user inspection through an interactive interface. In this case, peak picking is performed using Global Spectral Deconvolution [29], and binding is assessed by comparing the integrated signal intensities within predefined spectral regions across the different spectra. Among freely available tools, CcpNmr AnalysisScreen [146] provides an integrated environment for fragment-screening workflows. The software allows users to load and organize spectra together with relevant metadata, perform semi-automatic quality control procedures to identify and discard spectra affected by phasing problems, solvent artefacts, or missing signals, and optimize ligand mixtures. In addition, the platform allows the construction of customizable analysis pipelines tailored to the most common NMR techniques employed in fragment-based drug discovery, and provides dedicated tools for hit analysis and verification. Alongside these ligand-observed 1D techniques, automated platforms such as MUNIN [147] and the PICASSO web server [148] facilitate the analysis of protein-observed 2D spectra by systematically identifying the chemical shift perturbations that occur upon ligand binding.

To the best of our knowledge, the pipeline developed in this work represents the first automated workflow specifically tailored to the analysis of ^{1}H photo-CIDNP spectra. One of the central processing challenges in these datasets is the difficulty of obtaining reliable phase correction. In particular, strong signals arising from matrix compounds can introduce baseline distortions that make optimal phasing difficult or even impossible. In some existing analysis frameworks, such as CcpNmr AnalysisScreen [146], spectra for which satisfactory phase correction cannot be achieved are typically discarded during the quality control stage. To address this limitation, our pipeline operates on magnitude-mode spectra (see next section), which enforce an absorptive signal character across the spectra and allow reliable downstream analysis even when conventional phasing procedures fail.

A further challenge in photo-CIDNP experiments is the need to correlate information across several spectra acquired under different experimental conditions. In addition to ligand-only and ligand–protein spectra, measurements may include several relaxation delays as well as standard NMR spectra used to identify the hyperpolarised resonances. These hyperpolarised signals serve as probes for ligand–protein interaction and must therefore be tracked consistently across all experimental conditions. To perform this matching step, the pipeline employs the Hungarian matching algorithm, which allows flexible pairing of peaks across spectra. This approach avoids the limitations of fixed integration regions, which may lead to unreliable results when peak positions shift due to experimental factors such as temperature changes induced by sample illumination, protein addition, or buffer variations.

The pipeline is fully automated and produces results at multiple levels of analysis. At the peak level, it reports peak positions, amplitudes, photo-CIDNP enhancement, and the corresponding quenching upon protein addition. At the signal level, quenching measures are computed and tentative assignments are suggested. Finally, at the molecular level, ligands are classified as hit or no hit through different scoring modalities, including C_{bind} and Q_{bind}, which can be evaluated either on peak amplitudes or on integrated signal intensities. The detailed peak-level output also allows the results to be exported and used in further, more specialized analyses. For example, it is possible to study an effective combination of C_{bind} and Q_{bind} in a single, more sensitive score.

The development and validation of the pipeline were carried out exclusively on experimental spectra rather than on simulated datasets. This choice made it possible to directly investigate the specific characteristics of photo-CIDNP data and to optimize the pipeline so that it remains robust in the presence of typical experimental distortions.

For hit identification, selecting the signal exhibiting the largest quenching as representative of the entire molecule proved to be the most effective strategy. A situation in which one signal exhibits strong quenching while others show weaker effects is consistent with the structural interpretation that the corresponding atom, or group of magnetically equivalent atoms, lies closer to the protein binding site. Averaging the quenching values across signals, as proposed in previous approaches [145], would tend to suppress this information and may lead to the loss of potential hit candidates.

When the binding threshold is set to 0.2, the pipeline recovers 91.7% of the annotated hits using C_{bind} and 88.2% using Q_{bind} (see Table 7.6). Although the recall is excellent, the precision remains limited, indicating that the fully automated procedure tends to overselect potential hits. In the context of fragment screening, however, this outcome can still be considered advantageous. The algorithm ensures that no relevant hit candidates are missed, while ligands classified as non-hits can be safely discarded. Subsequent visual inspection can then be limited to a reduced subset of the original screening library, thereby accelerating the overall analysis workflow.

Beyond hit detection, the quenching measures computed at the signal level can be projected onto the ligand structure to obtain a qualitative indication of the binding epitope. A more quantitative characterization of the interaction would require independent measurements of binding affinities. The structural projection is obtained through a Bayesian assignment procedure that matches predicted chemical shifts, generated by a graph neural network trained on the NMRShiftDB dataset, with the experimental shifts measured in the photo-CIDNP spectra. A similar Bayesian matching strategy has previously been applied in the context of solid-state NMR [70].

In contrast to the one-to-one matching scheme used in that work, where each predicted shift is assigned to exactly one experimental signal, the present application requires a more flexible formulation. Only the hyperpolarised peaks are relevant for the binding analysis, and therefore both predicted and experimental shifts must be allowed to remain unmatched. Not all atoms in the molecule are sensitive to the CIDNP effect, and some experimental peaks may correspond to artefacts or signals unrelated to the ligand of interest. This bidirectional matching framework reduces the number of experimental peaks that must be considered, typically fewer than five hyperpolarised signals are observed in a spectrum, often only one or two, thereby reducing the computational cost of exploring the assignment space. At the same time, the problem becomes more challenging because the assignment relies primarily on the proximity between predicted and measured chemical shifts.

The difference between predicted and experimental shifts arises from two main sources: the intrinsic prediction error of the graph neural network and systematic shifts caused by temperature changes during sample illumination. The latter effect is partially corrected through a spectral alignment preprocessing step, although small residual shifts may remain in difficult cases. Nevertheless, the assignment performance (Figure 7.8) shows that in approximately 70% of the cases the most probable assignment corresponds to the correct one, a result consistent with the performance reported for analogous Bayesian assignment methods in one-dimensional ^{13}C solid-state NMR [70]. An improvement of the assignment algorithm can be achieved by including in the matching prior additional information on the multiplicity and coupling of the signals.

8.6 Magnitude Spectrum as a Strategy for Phase-Independent Analysis

Phasing a photo-CIDNP spectrum is a non-trivial task. The difficulty arises first from the simultaneous presence of absorptive (positive) and emissive (negative) resonances. Several widely used automatic phase correction methods rely on assumptions that are violated under these conditions. For instance, the ACME algorithm [76] incorporates a penalisation for negative peaks in the objective function in order to favour absorptive line shapes. While this assumption is appropriate for conventional liquid-state NMR spectra, it is not valid for photo-CIDNP data, where negative peaks are physically meaningful and correspond to emissive hyperpolarised signals. Consequently, enforcing positivity may lead to systematic phase distortions.

In this context the appropriate phasing criterion is not the suppression of negative peaks but rather the recovery of symmetric line shapes for all resonances and the suppression of dispersion patterns. To address this issue we implemented a cost function that simultaneously minimises the spectral entropy of the real part and enforces a vanishing integral of the imaginary component. This procedure yielded improved phasing performance compared with existing automatic approaches on the CIDNP dataset (Fig. 7.3), which consists of ligand-only spectra acquired at relatively high concentrations.

However, in fragment-based screening experiments the ligand concentration is significantly lower and the protein target is present in the sample. Under these conditions it was not possible to produce satisfactory phase correction even with phase parameters retrieved manually. This limitation represents a critical issue for automated analysis because the screening strategy relies on correctly correlating signals across multiple spectra. If different spectra are affected by residual phase distortions, the apparent peak positions, amplitudes, and line shapes may vary in a way that hampers the signal quenching detection. Ideally, all spectra would exhibit symmetric peaks with identical phase characteristics. Since this condition could not be guaranteed in practice, we adopted an alternative strategy and performed the analysis in magnitude mode in order to remove the phase degree of freedom. By construction, the magnitude spectrum is invariant with respect to such phase distortions, thereby providing a representation that is more stable for automated peak detection and cross-spectrum comparison.

The use of magnitude spectra for pattern recognition tasks in NMR spectra has previously been advocated by de B. Harrington and Wang [69]. In their work the magnitude representation was shown to be advantageous for predictive modelling, despite the appearance of asymmetric peak shapes.

We observed that part of this asymmetry originates from baseline distortions that become amplified by the non-linear magnitude transformation. To mitigate this effect we propose a processing protocol in which the real and imaginary components of the spectrum are independently baseline corrected before computing the magnitude spectrum. This additional step reduces distortions and produces more symmetric peak profiles (see Figure 7.4).

The usage of the magnitude spectrum is not standard practice in solution NMR. The reasons for the limited use of magnitude spectra in solution NMR are linked to the non-linearity introduced by the modulus operation that combines the real and imaginary components of the spectrum. This transformation modifies several spectral properties, including peak shape and noise statistics. In particular, the Gaussian noise distribution of

the complex spectrum becomes Rician in the magnitude representation. This effect must be taken into account when estimating the noise level, which is essential for detecting hyperpolarised signals.

Related considerations are well known in magnetic resonance imaging (MRI) [68], where magnitude images are routinely employed both for qualitative structural assessment (for instance in T_1- and T_2-weighted imaging) and in quantitative relaxometry mapping. Although the signal statistics differ from those of complex data, magnitude representations remain widely used because of their robustness to phase variations.

In the context of the present work, magnitude-mode processing improved the detection of hyperpolarised signals across all transverse relaxation delays acquired in the screening experiments, including conditions in which the overall signal-to-noise ratio decreases (Table 7.3). The benefit is particularly evident in the precision of the detection, with a substantial reduction of spurious peaks. These artefacts often originate from negative lobes produced by residual phase distortions in the absorptive representation. The magnitude transformation suppresses such dispersion-like patterns and produces symmetric peak shapes, thereby stabilising the peak detection procedure.

Following hyperpolarised signal detection, the algorithm identifies potential binding events by evaluating the quenching of ligand resonances upon addition of the target protein. This step relies on comparing peak amplitudes between the ligand-only and ligand–protein spectra. When performing this comparison in magnitude mode it is necessary to consider the line-shape modifications introduced by the modulus operation. The non-linear combination of the real and imaginary components results in increased linewidth and broader spectral tails.

Despite these distortions, the screening results indicate that amplitude ratios computed from magnitude spectra remain informative for the classification task. In particular, the hit/no-hit screening analysis (Table 7.6) successfully recovered all relevant binding candidates when peak amplitudes were evaluated in magnitude mode. At the same time, the regression metrics associated with the C_{bind} and Q_{bind} estimators exhibit non-negligible RMSE and MAE values. These discrepancies may arise from biases in amplitude estimation introduced by the magnitude transformation or from inaccuracies accumulated in earlier stages of the analysis pipeline.

A preliminary visual inspection suggests that the amplitude ratios themselves remain relatively robust with respect to the magnitude transformation, particularly because the hyperpolarised peaks in these experiments do not substantially overlap. Nevertheless, a more systematic investigation would be required to fully characterise the quantitative behaviour of magnitude-mode amplitudes in fragment-based drug discovery applications. Such an assessment would clarify the limits of this representation while further evaluating its principal advantage in the present context, namely the possibility of bypassing the challenging phase-correction problem encountered in photo-CIDNP screening experiments.

8.7 Toward Quantitative Binding Characterization

The development of the automatic deep learning based pipeline for fragment screening with ^{1}H NMR spectra served multiple purposes. In light of the recent introduction of photo-CIDNP NMR in drug discovery applications [11, 12], the design of such a pipeline provided an opportunity to better understand the structure of the data, the preprocess-

ing steps required, the potential pitfalls of the analysis, and the regularities that can be exploited for automated detection. At the same time, the implementation of the workflow clarifies which artefacts may arise in practice and how they propagate through the analysis. The resulting framework can be naturally integrated into a graphical interface in which automatic predictions are complemented by interactive tools, allowing the user to inspect and, if necessary, manually refine the output. In this way the system can assist the scientist in routine fragment screening while preserving the possibility of expert intervention.

At present, the binding detection stage is highly robust and reliably filters out ligands that do not exhibit detectable interaction with the target. The quantitative estimation of signal quenching, however, still requires further refinement. A crucial aspect of this improvement process is the identification of the origin of the observed inaccuracies. These may arise either from the accumulation of small errors introduced at earlier stages of the pipeline or from potential biases in amplitude estimation associated with the magnitude spectrum representation. Understanding the relative importance of these effects will allow the design of targeted corrections. In addition, as already discussed in the context of multiplet segmentation, the predictions should be accompanied by explicit uncertainty estimates in order to support decision making in the screening process. Such information would enable a reliable ranking of candidate binders, thereby restricting manual inspection to a smaller subset of highly promising molecules.

The insights obtained from the sequential pipeline naturally motivate the development of an end-to-end deep learning framework capable of addressing the task in a more integrated manner. In the current pipeline each processing stage receives a specific input, produces an intermediate output, and passes information to the next step. As a consequence, errors introduced in earlier stages may accumulate and propagate through the workflow. In contrast, an end-to-end architecture could jointly process all relevant inputs and infer the final prediction directly from the raw or minimally processed spectral data. The knowledge gained from the analysis of the individual steps can be incorporated into the design of the training data and the model architecture, thereby combining interpretability with improved predictive performance.

The training dataset for such a model could be generated synthetically by simulating spectra that reproduce realistic signal patterns of small molecules. The photo-CIDNP effect and ligand binding could be represented by selectively enhancing or attenuating the corresponding resonances. A particularly appealing research direction would be the development of a multi-modal architecture capable of learning simultaneously from one-dimensional NMR spectra and from molecular graph representations of the ligands. In such a framework the quenching measured in the spectrum could be mapped to the individual hydrogen atoms of the molecule, enabling a more detailed interpretation of the binding process.

A further extension towards quantitative analysis would involve the integration of spectra acquired at multiple ligand concentrations, as in titration experiments (see Figure 8.1). By measuring atom-wise signal quenching as a function of ligand concentration it becomes possible to extract the CIDNP dissociation constant (CIDNP-K_D) through appropriate fitting procedures [12]. Incorporating this information into the automated framework would therefore extend the methodology from qualitative screening towards quantitative characterization of binding affinity.

Another promising direction concerns the analysis of ligand mixtures. In this strategy multiple fragments are screened simultaneously in the presence of the target pro-

tein, reducing both the consumption of protein and the number of NMR experiments required. Moreover, mixture experiments naturally enable competition assays between ligands. The main challenge of this approach lies in the increased spectral congestion, which complicates the automatic interpretation, particularly in ^{1}H NMR where scalar couplings produce multiplet structures. An additional difficulty is the assignment of detected resonances to the corresponding ligands based on reference spectra. This task becomes considerably simpler in the case of ^{19}F NMR [149], where each ligand typically contributes a single singlet because only one fluorine atom, or a magnetically equivalent group of fluorine atoms, is present. Under these conditions the analysis could potentially be addressed using a framework similar to the one developed in this work.

Finally, an additional challenge is represented by low magnetic field regime. Benchtop NMR spectrometers are becoming increasingly attractive due to their accessibility and reduced operational requirements. In photo-CIDNP experiments this regime is particularly interesting because the enhancement effect is inversely proportional to the magnetic field strength, leading to stronger signal amplification. This advantage is counterbalanced by the reduction in spectral resolution and the increased degree of peak overlap, which complicate spectral interpretation. Developing automated analysis tools capable of operating reliably under these conditions therefore represents an important direction for future research.

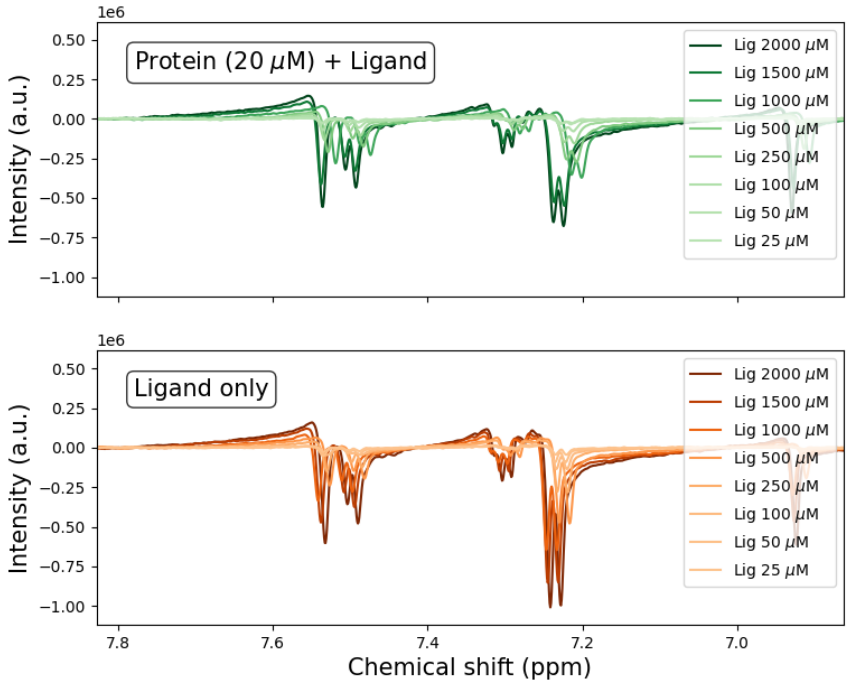


Figure 8.1: Photo-CIDNP spectra of ligand in the presence of the PGK1 protein (top) and of the ligand alone (bottom), recorded at different ligand concentrations.

Chapter 9

Conclusions

During the doctoral research presented in this thesis, we developed deep learning–based algorithms aimed at automating the interpretation of ^{1}H NMR spectra. Despite the maturity and widespread use of NMR spectroscopy in chemistry and structural biology, several stages of spectral analysis still rely on time–consuming manual annotation performed by trained experts, which does not scale well with the high throughput at which spectra can now be acquired.

Two complementary aspects of ^{1}H NMR data analysis were considered. The first concerns the characterization of the internal splitting pattern of resonances, which reflects scalar spin–spin couplings and provides information about the connectivity of the underlying spin system. The second concerns the comparison of spectra recorded under different experimental conditions in order to detect spectral changes associated with molecular interactions. In particular, the work focused on the identification of signal quenching effects in photo-CIDNP experiments used for ligand–protein binding detection.

Firstly we investigated the automatic classification of multiplet structures in ^{1}H NMR spectra. We formulated the problem as a segmentation task on one-dimensional spectral signals and explored convolutional neural networks as suitable models for this purpose. The resulting prototype architecture demonstrated that deep learning can capture the characteristic patterns produced by scalar couplings directly from spectral data, without relying on explicit parametric models of the lineshape. Building on this idea, a probabilistic deep learning framework, MuSe Net, was subsequently developed in order to improve both predictive accuracy and interpretability. In addition to predicting the most likely splitting class, the model provides a probability distribution over possible patterns, enabling a principled quantification of predictive uncertainty. This information can be exploited in downstream analysis pipelines, for instance by providing prior information to Bayesian spectral fitting procedures.

While these results demonstrate the potential of deep learning for multiplet classification, several challenges remain. The current models operate primarily on spectral phenotypes and therefore do not fully exploit global information about the spin system. One promising direction is the integration of MuSe Net predictions with physics–based spectral fitting procedures. In particular, the probabilistic output of the model could be incorporated as informative priors within Bayesian spectral fitting frameworks, combining the pattern recognition capabilities of neural networks with the interpretability of physical spectral models. A complementary direction concerns the development of models that explicitly incorporate molecular structure information. This could be achieved by generating synthetic spectra consistent with the molecular structure and coupling topol-

ogy, allowing the model to learn spectral patterns under physically plausible constraints and enabling a more global and chemically consistent interpretation of the spin system.

The second part of our work addressed the automation of ligand–protein interaction detection using ^{1}H photo-CIDNP NMR data. Photo-CIDNP experiments exploit hyperpolarization mechanisms to enhance specific resonances and are therefore employed in fragment-based screening experiments. The interpretation of these spectra, however, is complicated by experimental artefacts and by the need to compare signals across multiple experimental conditions. To address these challenges, we developed an automated analysis pipeline specifically tailored to photo-CIDNP datasets. The workflow combines spectral preprocessing, automatic peak detection, signal tracking across experiments, and the evaluation of signal quenching upon ligand binding. The resulting system enables the automatic identification of candidate binding fragments and the filtering of inactive compounds in large screening datasets. Validation on a real fragment screening campaign showed that the pipeline can reliably detect binding events while substantially reducing the manual effort required for spectral inspection, thereby supporting the prioritization of fragments for further experimental investigation.

For photo-CIDNP analysis, an important direction for future work concerns the quantitative estimation of binding strength. The present pipeline focuses primarily on the reliable identification of binding events, while the quantification of signal quenching remains affected by systematic uncertainties arising from spectral distortions and experimental variability. Understanding the origin of these biases and developing appropriate correction strategies will be necessary in order to enable reliable ranking of ligand candidates based on NMR data alone. In addition, the integration of explicit uncertainty estimates in the prediction pipeline would provide valuable information for guiding experimental decision making.

Several methodological extensions can also be envisaged. One promising direction concerns the analysis of ligand mixtures, both with ^{1}H and ^{19}F isotopes, which would enable competition experiments and significantly increase the throughput of fragment screening campaigns. Another important challenge is the development of algorithms capable of operating reliably in low-field NMR regimes, such as those encountered with benchtop spectrometers. In these conditions the reduced spectral resolution leads to stronger peak overlap, making automated interpretation substantially more difficult. Addressing these challenges will likely require the combination of improved signal processing techniques with learning methods capable of exploiting prior structural information.

The research presented here was carried out in the context of two industrial collaborations supported by Innosuisse. The partners were Bruker AG, a leading manufacturer of scientific instrumentation and a major producer of high–resolution NMR spectrometers, and NexMR AG, an ETH Zurich spin–off developing photo-CIDNP NMR methods for fragment-based drug discovery. These collaborations provided access to experimental datasets and domain expertise and ensured that the methodological developments addressed practical challenges encountered in real spectroscopic workflows.

The approaches developed in our work show that data–driven models can be integrated with established NMR knowledge in order to assist the interpretation of complex spectra and the identification of ligand–protein interactions in fragment screening experiments. Complementary to traditional processing procedures, the proposed methods provide a framework in which learning-based models and spectroscopic reasoning can operate jointly within automated analysis pipelines to support decision making in chemical research.

Bibliography

[1] Otto Robert Frisch and Otto Stern. "Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. I". In: *Zeitschrift für Physik* 85.1–2 (1933), pp. 4–16. DOI: 10.1007/BF01330773.

[2] Isidor Isaac Rabi, Jerrold Reinach Zacharias, Sidney Millman, and Polykarp Kusch. "A New Method of Measuring Nuclear Magnetic Moment". In: *Physical Review* 53 (1938), p. 318. DOI: 10.1103/PhysRev.53.318.

[3] Edward Mills Purcell, Henry Cutler Torrey, and Robert Vivian Pound. "Resonance Absorption by Nuclear Magnetic Moments in a Solid". In: *Physical Review* 69 (1946), pp. 37–38. DOI: 10.1103/PhysRev.69.37.

[4] Felix Bloch, William Webster Hansen, and Martin Packard. "Nuclear Induction". In: *Physical Review* 69 (1946), p. 127. DOI: 10.1103/PhysRev.69.127.

[5] Richard R Ernst and Weston A Anderson. "Application of Fourier transform spectroscopy to magnetic resonance". In: *Review of Scientific Instruments* 37.1 (1966), pp. 93–102. DOI: 10.1063/1.1719961.

[6] Yong-Cheng Ning. *Interpretation of Organic Spectra*. John Wiley & Sons, 2011. ISBN: 9780470825167. DOI: 10.1002/9780470825181.

[7] Timothy D. W. Claridge. *High-Resolution NMR Techniques in Organic Chemistry*. 3rd. Amsterdam: Elsevier Science, 2016. ISBN: 9780080999869. DOI: 10.1016/C2015-0-04654-8.

[8] Alessia Vignoli, Veronica Ghini, Gaia Meoni, Cristina Licari, Panteleimon G. Takis, Leonardo Tenori, Paola Turano, and Claudio Luchinat. "High-Throughput Metabolomics by 1D NMR". In: *Angewandte Chemie International Edition* 58.4 (2019), pp. 968–994. DOI: 10.1002/anie.201804736.

[9] Teresa W.-M. Fan and Andrew N. Lane. "Applications of NMR Spectroscopy to Systems Biochemistry". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 92-93 (2016), pp. 18–53. DOI: 10.1016/j.pnmrs.2016.01.005.

[10] Maurizio Pellecchia, Daniel S. Sem, and Kurt Wüthrich. "NMR in drug discovery". In: *Nature Reviews Drug Discovery* 1.3 (2002), pp. 211–219. DOI: 10.1038/nrd748.

[11] Felix Torres, Matthias Bütikofer, Gabriela R. Stadler, Alois Renn, Harindranath Kadavath, Raitis Bobrovs, Kristaps Jaudzems, and Roland Riek. "Ultrafast Fragment Screening Using Photo-Hyperpolarized (CIDNP) NMR". In: *Journal of the American Chemical Society* 145.22 (2023), pp. 12066–12080. DOI: 10.1021/jacs.3c01392.

[12] Matthias Bütikofer, Gabriela R. Stadler, Harindranath Kadavath, Riccardo Cadalbert, Felix Torres, and Roland Riek. "Rapid Protein–Ligand Affinity Determination by Photoinduced Hyperpolarized NMR". In: *Journal of the American Chemical Society* 146.26 (2024), pp. 17974–17985. DOI: 10.1021/jacs.4c04000.

[13] Nils Lorz, Barbara Czarniecki, Sandra Loss, Benno Meier, and Alvar D. Gossert. "Higher Contrast in ^{1}H-Observed NMR Ligand Screening with the PEARLScreen Experiment". In: *Angewandte Chemie International Edition* 64.17 (2025), e202423879. DOI: 10.1002/anie.202423879.

[14] Komal Zia, Talal Siddiqui, Saqib Ali, Imran Farooq, Muhammad Sohail Zafar, and Zohaib Khurshid. "Nuclear Magnetic Resonance Spectroscopy for Medical and Dental Applications: A Comprehensive Review". In: *European Journal of Dentistry* 13.1 (2019), pp. 124–128. DOI: 10.1055/s-0039-1688654.

[15] Simon Wiemers-Meyer, Martin Winter, and Sascha Nowak. "NMR as a Powerful Tool to Study Lithium-Ion Battery Electrolytes". In: *Annual Reports on NMR Spectroscopy*. Ed. by David M. Grant. Vol. 97. Annual Reports on NMR Spectroscopy. Elsevier, 2019, pp. 121–162. DOI: 10.1016/bs.arnmr.2018.12.003.

[16] L. A. Cardoza, A. K. Korir, W. H. Otto, C. J. Wurrey, and C. K. Larive. "Applications of NMR Spectroscopy in Environmental Science". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 45.3-4 (2004), pp. 209–238. DOI: 10.1016/j.pnmrs.2004.06.002.

[17] Andre J. Simpson, Myrna J. Simpson, and Ronald Soong. "Environmental Nuclear Magnetic Resonance Spectroscopy: An Overview and a Primer". In: *Analytical Chemistry* 90.1 (2018), pp. 628–639. DOI: 10.1021/acs.analchem.7b03241.

[18] Veronica Lolli and Augusta Caligiani. "How nuclear magnetic resonance contributes to food authentication: current trends and perspectives". In: *Current Opinion in Food Science* 58 (2024), p. 101200. DOI: 10.1016/j.cofs.2024.101200.

[19] Pierluigi Mazzei and Alessandro Piccolo. "HRMAS NMR spectroscopy applications in agriculture". In: *Chemical and Biological Technologies in Agriculture* 4.1 (2017), p. 11. DOI: 10.1186/s40538-017-0093-9.

[20] Ruge Cao, Xinru Liu, Yuqian Liu, Xuqing Zhai, Tianya Cao, Aili Wang, and Ju Qiu. "Applications of nuclear magnetic resonance spectroscopy to the evaluation of complex food constituents". In: *Food Chemistry* 342 (2021), p. 128258. DOI: 10.1016/j.foodchem.2020.128258.

[21] J. Keeler. *Understanding NMR Spectroscopy*. Wiley, 2011. ISBN: 9781119964933.

[22] Roman Koradi, Markus Billeter, Martin Engeli, Peter Güntert, and Kurt Wüthrich. "Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY". In: *Journal of Magnetic Resonance* 135.2 (1998), pp. 288–297. ISSN: 1090-7807. DOI: 10.1006/jmre.1998.1570.

[23] S. Boentges. "Local Symmetry in 2D and 3D NMR Spectra". In: *Journal of Magnetic Resonance* 85.2 (1989), pp. 219–230. DOI: 10.1016/0022-2364(89)90148-0.

[24] Suhas Tikole, Victor Jaravine, Vladimir Rogov, Volker Dötsch, and Peter Güntert. "Peak picking NMR spectral data using non-negative matrix factorization". In: *BMC Bioinformatics* 15 (2014), p. 46. DOI: 10.1186/1471-2105-15-46.

[25] Babak Alipanahi, Xin Gao, Emre Karakoc, Logan Donaldson, and Ming Li. "PICKY: a novel SVD-based NMR spectra peak picking method". In: *Bioinformatics* 25.12 (May 2009), pp. i268–i275. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btp225.

[26] M. A. Delsuc and G. C. Levy. "The application of maximum entropy processing to the deconvolution of coupling patterns in NMR". In: *Journal of Magnetic Resonance* 76.2 (1988), pp. 306–315. DOI: 10.1016/0022-2364(88)90112-6.

[27] M. J. Seddon. "Automatic Recognition of Multiplet Patterns and Measurement of Coupling Constants in NMR and EPR Spectra through the Application of Maximum Entropy". In: *Journal of Magnetic Resonance* 119.1 (1996), pp. 72–83. DOI: 10.1006/jmra.1996.0072.

[28] V. Stoven, J. P. Annereau, M. A. Delsuc, and J. Y. Lallemand. "A New N-Channel Maximum Entropy Method in NMR for Automatic Reconstruction of "Decoupled Spectra" and J-Coupling Determination". In: *Journal of Chemical Information and Computer Sciences* 37.2 (1997), pp. 265–272. DOI: 10.1021/ci960321n.

[29] Carlos Cobas, Felipe Seoane, Santiago Domínguez, Stan Sykora, and Antony N. Davies. "A new approach to improving automated analysis of proton NMR spectra through Global Spectral Deconvolution (GSD)". In: *Spectroscopy Europe* 23.1 (2011), pp. 26–30.

[30] Yichen Cheng, Xin Gao, and Faming Liang. "Bayesian Peak Picking for NMR Spectra". In: *Genomics, Proteomics & Bioinformatics* 12.1 (2014), pp. 39–47. ISSN: 1672-0229. DOI: 10.1016/j.gpb.2013.07.003.

[31] Charles Buchanan, Gogulan Karunanithy, Olga Tkachenko, Michael Barber, Michael T. Marty, Timothy J. Nott, Christina Redfield, and Andrew J. Baldwin. "UnidecNMR: automatic peak detection for NMR spectra in 1–4 dimensions". In: *Nature Communications* 16 (2025), p. 449. DOI: 10.1038/s41467-024-54899-3.

[32] Dominique Jeannerat and Geoffrey Bodenhausen. "Determination of Coupling Constants by Deconvolution of Multiplets in NMR". In: *Journal of Magnetic Resonance* 141.1 (1999), pp. 133–140. DOI: 10.1006/jmre.1999.1845.

[33] Damien Jeannerat and Carlos Cobas. "Application of multiplet structure deconvolution to extract scalar coupling constants from 1D nuclear magnetic resonance spectra". In: *Magnetic Resonance* 2 (2021), pp. 545–555. DOI: 10.5194/mr-2-545-2021.

[34] S. S. Golotvin and V. A. Chertkov. "Pattern Recognition of the Multiplet Structure of NMR Spectra". In: *Russian Chemical Bulletin* 46 (1997), pp. 423–430. DOI: 10.1007/BF02495389.

[35] Thomas R. Hoye, Paul R. Hanson, and James R. Vyvyan. "A Practical Guide to First-Order Multiplet Analysis in ^{1}H NMR Spectroscopy". In: *The Journal of Organic Chemistry* 59.15 (1994), pp. 4096–4103. DOI: 10.1021/jo00094a018.

[36] Thomas R. Hoye and Hongyu Zhao. "A Method for Easily Determining Coupling Constant Values: An Addendum to "A Practical Guide to First-Order Multiplet Analysis in ^{1}H NMR Spectroscopy"". In: *The Journal of Organic Chemistry* 67.12 (2002), pp. 4014–4016. DOI: 10.1021/jo001139v.

[37] Sergey Golotvin, Eugene Vodopianov, and Antony J. Williams. "A new approach to automated first-order multiplet analysis". In: *Magnetic Resonance in Chemistry* 40.5 (2002), pp. 331–336. DOI: 10.1002/mrc.1014.

[38] J. C. Cobas, V. Constantino-Castillo, M. Martín-Pastor, and F. del Río-Portilla. "A two-stage approach to automatic determination of ^{1}H NMR coupling constants". In: *Magnetic Resonance in Chemistry* 43.10 (2005), pp. 843–848. DOI: 10.1002/mrc.1623.

[39] Lyn Griffiths. "Towards the automatic analysis of ^{1}H NMR spectra". In: *Magnetic Resonance in Chemistry* 38.6 (2000), pp. 444–452. DOI: 10.1002/1097-458X(200006)38:6<444::AID-MRC673>3.0.CO;2-Z.

[40] AlphaFold Protein Structure Database. *AlphaFold Protein Structure Database*. https://alphafold.ebi.ac.uk/. Developed by DeepMind and EMBL-EBI. 2021.

[41] Dicheng Chen, Zi Wang, Di Guo, Vladislav Orekhov, Yulan Lin, and Xiaohua Zou. "Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy". In: *Chemistry - A European Journal* 26.46 (2020), pp. 10391–10401. DOI: 10.1002/chem.202000246.

[42] Juan C. Cobas, Santiago Gil Baguena, Fernando Seoane, Stefan Sykora, Cristina Garcia Fernandez, and Hermenegildo Domínguez. "NMR signal processing, prediction, and structure verification with machine learning techniques". In: *Magnetic Resonance in Chemistry* 58.6 (2020), pp. 512–519. DOI: 10.1002/mrc.4989.

[43] Amir Jahangiri and Vladislav Orekhov. "Beyond Traditional Magnetic Resonance Processing with Artificial Intelligence". In: *Communications Chemistry* 7 (2024), p. 244. DOI: 10.1038/s42004-024-01325-w.

[44] Piotr Klukowski, Roland Riek, and Peter Güntert. "Machine learning in NMR spectroscopy". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 148–149 (2025), p. 101575. DOI: 10.1016/j.pnmrs.2025.101575.

[45] Piotr Klukowski, Maciej Drwal, Adam Gonczarek, and Michał J. Walczak. "NMR-Net: a deep learning approach to automated peak picking of protein NMR spectra". In: *Bioinformatics* 34.15 (2018), pp. 2590–2597. DOI: 10.1093/bioinformatics/bty134.

[46] Da-Wei Li, Alexandar L. Hansen, Chunhua Yuan, Lei Bruschweiler-Li, and Rafael Brüschweiler. "DEEP Picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra". In: *Nature Communications* 12.1 (2021), p. 5229. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25496-5.

[47] Piotr Klukowski, Roland Riek, and Peter Güntert. "Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA". In: *Nature Communications* 13 (2022), p. 6151. DOI: 10.1038/s41467-022-33879-5.

[48] Amir Jahangiri, Tatiana Agback, Ulrika Brath, and Vladislav Orekhov. *Towards Ultimate NMR Resolution with Deep Learning*. 2025. DOI: 10.48550/arXiv.2502.20793. arXiv: 2502.20793 [physics.bio-ph].

[49] N. Schmid, S. Bruderer, F. Paruzzo, G. Fischetti, G. Toscano, D. Graf, M. Fey, A. Henrici, V. Ziebart, B. Heitmann, H. Grabner, J. D. Wegner, R. K. O. Sigel, and D. Wilhelm. "Deconvolution of 1D NMR spectra: A deep learning-based approach". In: *Journal of Magnetic Resonance* 347 (2023), p. 107357. ISSN: 1090-7807. DOI: 10.1016/j.jmr.2022.107357.

[50] Kaina Shibata and Hiromasa Kaneko. "Prediction of spin–spin coupling constants with machine learning in NMR". In: *Analytical Science Advances* 2.9–10 (2021), pp. 464–469. DOI: 10.1002/ansa.202000180.

[51] Jia Fang, Linyuan Hu, Jianfeng Dong, Haowei Li, Hui Wang, Huafen Zhao, Yao Zhang, and Min Liu. "Predicting scalar coupling constants by graph angle-attention neural network". In: *Scientific Reports* 11 (2021), p. 18686. DOI: 10.1038/s41598-021-97146-1.

[52] Giulia Fischetti, Nicolas Schmid, Simon Bruderer, Guido Caldarelli, Alessandro Scarso, Andreas Henrici, and Dirk Wilhelm. "Automatic classification of signal regions in 1H Nuclear Magnetic Resonance spectra". In: *Frontiers in Artificial Intelligence* 5 (2023). ISSN: 2624-8212. DOI: 10.3389/frai.2022.1116416.

[53] Tara N. Sainath, Oriol Vinyals, Andrew Senior, and Haşim Sak. "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks". In: *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015, pp. 4580–4584. DOI: 10.1109/ICASSP.2015.7178838.

[54] Ronald Mutegeki and Dong Seog Han. "A CNN-LSTM Approach to Human Activity Recognition". In: *Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. 2020, pp. 362–366. DOI: 10.1109/ICAIIC48513.2020.9065078.

[55] Gaowei Xu, Tianhe Ren, Yu Chen, and Wenliang Che. "A One-Dimensional CNN-LSTM Model for Epileptic Seizure Recognition Using EEG Signal Analysis". In: *Frontiers in Neuroscience* 14 (2020), p. 578126. DOI: 10.3389/fnins.2020.578126.

[56] Abdulkadir Tasdelen and Baha Sen. "A hybrid CNN-LSTM model for pre-miRNA classification". In: *Scientific Reports* 11 (2021), p. 14125. DOI: 10.1038/s41598-021-93656-0.

[57] Fatma Özgë Özkök and Mete Celik. "A hybrid CNN-LSTM model for high resolution melting curve classification". In: *Biomedical Signal Processing and Control* 71 (2022), p. 103168. DOI: 10.1016/j.bspc.2021.103168.

[58] Emily Sullivan. "Understanding from Machine Learning Models". In: *The British Journal for the Philosophy of Science* 73.1 (2022), pp. 109–133. DOI: 10.1093/bjps/axz035.

[59] Thomas Grote, Konstantin Genin, and Emily Sullivan. "Reliability in Machine Learning". In: *Philosophy Compass* 19.5 (2024), e12974. DOI: 10.1111/phc3.12974.

[60] Michael Mayer and Birgit Meyer. "Characterization of ligand binding by saturation transfer difference NMR spectroscopy". In: *Angewandte Chemie International Edition* 38.12 (1999), pp. 1784–1788. DOI: 10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.

[61] Claudio Dalvit, Paul Pevarello, Marco Tatò, Marco Veronesi, Anna Vulpetti, and Magnus Sundström. "Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water". In: *Journal of Biomolecular NMR* 18.1 (2000), pp. 65–68. DOI: 10.1023/a:1008354229396.

[62] Philip J. Hajduk, Edward T. Olejniczak, and Stephen W. Fesik. "One-Dimensional Relaxation- and Diffusion-Edited NMR Methods for Screening Compounds That Bind to Macromolecules". In: *Journal of the American Chemical Society* 119.50 (1997), pp. 12257–12261. DOI: 10.1021/ja9715962.

[63] J. H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning, and K. Golman. "Increase in signal-to-noise ratio of >10 000 times in liquid-state NMR". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.18 (2003), pp. 10158–10163. DOI: 10.1073/pnas.1733835100.

[64] C. Russell Bowers and Daniel P. Weitekamp. "Transformation of Symmetrization Order to Nuclear-Spin Magnetization by Chemical Reaction and Nuclear Magnetic Resonance". In: *Physical Review Letters* 57.21 (1986), pp. 2645–2648. DOI: 10.1103/PhysRevLett.57.2645.

[65] Ralph W. Adams, Juan A. Aguilar, Kevin D. Atkinson, Michael J. Cowley, Paul I. P. Elliott, Simon B. Duckett, Gary G. R. Green, Iman G. Khazal, Joaquín López-Serrano, and David C. Williamson. "Reversible Interactions with para-Hydrogen Enhance NMR Sensitivity by Polarization Transfer". In: *Science* 323.5922 (2009), pp. 1708–1711. DOI: 10.1126/science.1168877.

[66] Michael Goez. "Photo-CIDNP Spectroscopy". In: *Annual Reports on NMR Spectroscopy*. Ed. by Graham A. Webb. Vol. 59. Annual Reports on NMR Spectroscopy. Elsevier, 2009, pp. 109–138. DOI: 10.1016/S0066-4103(08)00403-1.

[67] Lars T. Kuhn. "Photo-CIDNP NMR spectroscopy of amino acids and proteins". In: *Topics in Current Chemistry* 338 (2013), pp. 229–300. DOI: 10.1007/128_2013_427.

[68] Hákon Gudbjartsson and Samuel Patz. "The Rician Distribution of Noisy MRI Data". In: *Magnetic Resonance in Medicine* 34.6 (1995), pp. 910–914. DOI: 10.1002/mrm.1910340618.

[69] Peter de B. Harrington and Xinyi Wang. "Spectral Representation of Proton NMR Spectroscopy for the Pattern Recognition of Complex Materials". In: *Journal of Analysis and Testing* 1.2 (2017), p. 10. ISSN: 2509-4696. DOI: 10.1007/s41664-017-0003-y.

[70] Manuel Córdova, Martins Balodis, Bruno Simões de Almeida, Michele Ceriotti, and Lyndon Emsley. "Bayesian probabilistic assignment of chemical shifts in organic solids". In: *Science Advances* 7.48 (2021), eabk2341. DOI: 10.1126/sciadv.abk2341.

[71] Jongmin Han, Hyungu Kang, Seokho Kang, Youngchun Kwon, Dongseon Lee, and Youn-Suk Choi. "Scalable graph neural network for NMR chemical shift prediction". In: *Physical Chemistry Chemical Physics* 24 (2022), pp. 26870–26878. DOI: 10.1039/D2CP04542G.

[72] Malcolm H. Levitt. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2nd ed. Chichester, UK: John Wiley & Sons, 2008. ISBN: 9780470511176.

[73] Richard R. Ernst. "Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture)". In: *Angewandte Chemie International Edition in English* 31.7 (1992), pp. 805–823. DOI: 10.1002/anie.199208053.

[74] J. A. Pople, W. G. Schneider, and H. J. Bernstein. "The Analysis of Nuclear Magnetic Resonance Spectra". In: *Canadian Journal of Chemistry* 35.8 (1957), pp. 1060–1072. DOI: 10.1139/v57-143.

[75] Richard R. Ernst, Geoffrey Bodenhausen, and Alexander Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Vol. 14. International Series of Monographs on Chemistry. Oxford, UK: Oxford University Press, 1987. ISBN: 9780198556473.

[76] Li Chen, Zhiqiang Weng, LaiYoong Goh, and Marc Garland. "An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization". In: *Journal of Magnetic Resonance* 158.1 (2002), pp. 164–168. ISSN: 1090-7807. DOI: 10.1016/S1090-7807(02)00069-1.

[77] Zhi Liu, Ahmed Abbas, Bing-Yu Jing, and Xin Gao. "WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering". In: *Bioinformatics* 28.7 (2012), pp. 914–920. DOI: 10.1093/bioinformatics/bts078.

[78] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. "WaveNet: A Generative Model for Raw Audio". In: *arXiv preprint arXiv:1609.03499* (2016).

[79] Zhouao Zhou, Xinli Liao, Xu Qiu, Yue Zhang, Jiyang Dong, Xiaobo Qu, and Donghai Lin. "NMRformer: A Transformer-Based Deep Learning Framework for Peak Assignment in 1D ^{1}H NMR Spectroscopy". In: *Analytical Chemistry* 97.1 (2025), pp. 904–911. DOI: 10.1021/acs.analchem.4c05632.

[80] Frank Hu, Michael S. Chen, Grant M. Rotskoff, Matthew W. Kanan, and Thomas E. Markland. "Accurate and Efficient Structure Elucidation from Routine One-Dimensional NMR Spectra Using Multitask Machine Learning". In: *ACS Central Science* 10.11 (2024), pp. 2162–2170. DOI: 10.1021/acscentsci.4c01132.

[81] Sai Krishna Devata et al. "DeepSPInN – deep reinforcement learning for molecular structure prediction from infrared and ^{13}C NMR spectra". In: *Digital Discovery* (2024). DOI: 10.1039/D4DD00008K.

[82] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. "Review the state-of-the-art technologies of semantic segmentation based on deep learning". In: *Neurocomputing* 493 (2022), pp. 626–646. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2022.01.005.

[83] Giulia Fischetti, Nicolas Schmid, Simon Bruderer, Björn Heitmann, Andreas Henrici, Alessandro Scarso, Guido Caldarelli, and Dirk Wilhelm. "A deep learning framework for multiplet splitting classification in 1H NMR". In: *Journal of Magnetic Resonance* 373 (2025), p. 107851. ISSN: 1090-7807. DOI: 10.1016/j.jmr.2025.107851.

[84] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going Deeper with Convolutions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.

[85] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal Loss for Dense Object Detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.2 (2020), pp. 318–327. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2018.2858826.

[86] Daochen Zha, Zaid Pervaiz Bhat, Kwei-Herng Lai, Fan Yang, Zhimeng Jiang, Shaochen Zhong, and Xia Hu. "Data-centric Artificial Intelligence: A Survey". In: *ACM Computing Surveys* 57.5 (2025), p. 129. ISSN: 0360-0300. DOI: 10.1145/3711118.

[87] John F. Kielkopf. "New approximation to the Voigt function with applications to spectral-line profile analysis". In: *J. Opt. Soc. Am.* 63.8 (1973), pp. 987–995. DOI: 10.1364/JOSA.63.000987.

[88] Favour Danladi Makurvet. "Biologics vs. Small Molecules: Drug Costs and Patient Access". In: *Medicine in Drug Discovery* 9 (2021), p. 100075. DOI: 10.1016/j.medidd.2020.100075.

[89] Rafael Padilla, Sérgio L. Netto, and Eduardo A. B. da Silva. "A Survey on Performance Metrics for Object-Detection Algorithms". In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. Niterói, Brazil, 2020, pp. 237–242. DOI: 10.1109/IWSSIP48289.2020.9145130.

[90] Mohammad Hossin and Mohd Najib Sulaiman. "A Review on Evaluation Metrics for Data Classification Evaluations". In: *International Journal of Data Mining & Knowledge Management Process* 5.2 (2015), pp. 1–11. DOI: 10.5121/ijdkp.2015.5201.

[91] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, Pablo Martinez-Gonzalez, and Jose Garcia-Rodriguez. "A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation". In: *Applied Soft Computing* 70 (2018), pp. 41–65. DOI: 10.1016/j.asoc.2018.05.018.

[92] Margherita Grandini, Enrico Bagli, and Giorgio Visani. *Metrics for Multi-Class Classification: an Overview*. 2020. arXiv: 2008.05756 [stat.ML].

[93] Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence, eds. *Dataset Shift in Machine Learning*. Neural Information Processing Series. Cambridge, MA: The MIT Press, 2009. ISBN: 9780262170055.

[94] Hyeong Hun Lee and Hyeonjin Kim. "Intact Metabolite Spectrum Mining by Deep Learning in Proton Magnetic Resonance Spectroscopy of the Brain". In: *Magnetic Resonance in Medicine* 82.1 (2019), pp. 33–48. DOI: 10.1002/mrm.27727.

[95] D. Flemming Hansen. "Using Deep Neural Networks to Reconstruct Non-uniformly Sampled NMR Spectra". In: *Journal of Biomolecular NMR* 73 (2019), pp. 577–585. DOI: 10.1007/s10858-019-00265-1.

[96] Xiaobo Qu, Yihui Huang, Hengfa Lu, Tianyu Qiu, Di Guo, Tatiana Agback, Vladislav Orekhov, and Zhong Chen. "Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning". In: *Angewandte Chemie International Edition* 59.26 (2020), pp. 10297–10300. DOI: 10.1002/anie.201908162.

[97] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. "On Calibration of Modern Neural Networks". In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*. Vol. 70. Proceedings of Machine Learning Research. Sydney, Australia: PMLR, 2017, pp. 1321–1330.

[98] Roman A. Valiulin. *NMR Multiplet Interpretation: An Infographic Walk-Through*. Berlin, Germany: De Gruyter, 2019, p. 166. ISBN: 9783110608403.

[99] Laurent Valentin Jospin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bennamoun. "Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users". In: *IEEE Computational Intelligence Magazine* 17.2 (2022), pp. 29–48. DOI: 10.1109/MCI.2022.3155327.

[100] Tim Salimans, Diederik P. Kingma, and Max Welling. "Markov Chain Monte Carlo and Variational Inference: Bridging the Gap". In: *Proceedings of the 32nd International Conference on Machine Learning (ICML)*. Vol. 37. JMLR Workshop and Conference Proceedings. Lille, France, 2015, pp. 1218–1226.

[101] Jeremiah Liu, Zi Lin, Shreyas Padhy, Dustin Tran, Tania Bedrax-Weiss, and Balaji Lakshminarayanan. "Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness". In: *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. Ed. by Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin. Curran Associates, Inc., 2020.

[102] Jeremiah Zhe Liu, Shreyas Padhy, Jie Ren, Zi Lin, Yeming Wen, Ghassen Jerfel, Zachary Nado, Jasper Snoek, Dustin Tran, and Balaji Lakshminarayanan. "A Simple Approach to Improve Single-Model Deep Uncertainty via Distance-Awareness". In: *Journal of Machine Learning Research* 24.42 (2023), pp. 1–63. URL: https://jmlr.org/papers/v24/22-0479.html.

[103] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of the 33rd International Conference on Machine Learning (ICML)*. Vol. 48. Proceedings of Machine Learning Research. New York, NY, USA: PMLR, 2016, pp. 1050–1059.

[104] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles". In: *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates, Inc., 2017.

[105] Abhijit Guha Roy, Jie Ren, Shekoofeh Azizi, Aaron Loh, Vivek Natarajan, Basil Mustafa, Nick Pawlowski, Jan Freyberg, Yuan Liu, Zach Beaver, Nam Vo, Peggy Bui, Samantha Winter, Patricia MacWilliams, Greg S. Corrado, Umesh Telang, Yun Liu, Taylan Cemgil, Alan Karthikesalingam, Balaji Lakshminarayanan, and Jim Winkens. "Does Your Dermatology Classifier Know What It Doesn't Know? Detecting the Long-Tail of Unseen Conditions". In: *Medical Image Analysis* 75 (2022), p. 102274. DOI: 10.1016/j.media.2021.102274.

[106] Mahesh Gour and Sweta Jain. "Uncertainty-Aware Convolutional Neural Network for COVID-19 X-Ray Images Classification". In: *Computers in Biology and Medicine* 140 (2022), p. 105047. DOI: 10.1016/j.compbiomed.2021.105047.

[107] Xiao Sun, Bahador Bahmani, Nikolaos N. Vlassis, WaiChing Sun, and Yanxun Xu. "Data-Driven Discovery of Interpretable Causal Relations for Deep Learning Material Laws with Uncertainty Propagation". In: *Granular Matter* 24 (2022), p. 1. DOI: 10.1007/s10035-021-01137-y.

[108] Mudasir Ahmad Ganaie, Minghui Hu, Ashwani Kumar Malik, Mohammad Tanveer, and Ponnuthurai Nagaratnam Suganthan. "Ensemble Deep Learning: A Review". In: *Engineering Applications of Artificial Intelligence* 115 (2022), p. 105151. DOI: 10.1016/j.engappai.2022.105151.

[109] Rahul Rahaman and Alexandre Thiery. "Uncertainty Quantification and Deep Ensembles". In: *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*. Curran Associates, Inc., 2021.

[110] Telmo Silva Filho, Hao Song, Miquel Perello-Nieto, Raul Santos-Rodriguez, Meelis Kull, and Peter Flach. "Classifier Calibration: A Survey on How to Assess and Improve Predicted Class Probabilities". In: *Machine Learning* 112.9 (2023), pp. 3211–3260. DOI: 10.1007/s10994-023-06336-7.

[111] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York, NY: Springer, 2009. ISBN: 9780387848570. DOI: 10.1007/978-0-387-84858-7.

[112] Y. Geeta Dharani, Nimisha G. Nair, Pallavi Satpathy, and Jabez Christopher. "Covariate Shift: A Review and Analysis on Classifiers". In: *2019 Global Conference for Advancement in Technology (GCAT)*. Bangalore, India: IEEE, 2019. DOI: 10.1109/GCAT47503.2019.8978471.

[113] José A. Montoya, Gudelia Figueroa P., and David González-Sánchez. "Statistical Inference for the Weitzman Overlapping Coefficient in a Family of Distributions". In: *Applied Mathematical Modelling* 71 (2019), pp. 558–568. DOI: 10.1016/j.apm.2019.02.036.

[114] Flavio De Lorenzi, Tom Weinmann, Simon Bruderer, Björn Heitmann, Andreas Henrici, and Simon Stingelin. "Bayesian analysis of 1D ^{1}H-NMR spectra". In: *Journal of Magnetic Resonance* 364 (2024), p. 107723. DOI: 10.1016/j.jmr.2024.107723.

[115] Karel Kouřil and Benno Meier. "Hyperpolarization and Sensitivity in Nuclear Magnetic Resonance". In: *Journal of Magnetic Resonance Open* 21 (2024), p. 100172. DOI: 10.1016/j.jmro.2024.100172.

[116] Joachim Bargon, Hanns Fischer, and Ulrich Johnsen. "Kernresonanz-Emissionslinien während rascher Radikalreaktionen: I. Aufnahmeverfahren und Beispiele". In: *Zeitschrift für Naturforschung A* 22.10 (1967), pp. 1551–1555. DOI: 10.1515/zna-1967-1014.

[117] Harold Roy Ward and Ronald G. Lawler. "Nuclear Magnetic Resonance Emission and Enhanced Absorption in Rapid Organometallic Reactions". In: *Journal of the American Chemical Society* 89.21 (1967), pp. 5518–5519. DOI: 10.1021/ja00997a078.

[118] Gerhard L. Closs. "A Mechanism Explaining Nuclear Spin Polarizations in Radical Combination Reactions". In: *Journal of the American Chemical Society* 91.16 (1969), pp. 4552–4554. DOI: 10.1021/ja01044a043.

[119] R. Kaptein and L. J. Oosterhoff. "Chemically Induced Dynamic Nuclear Polarization III: Anomalous Multiplets of Radical Coupling and Disproportionation Products". In: *Chemical Physics Letters* 4.4 (1969), pp. 214–216. DOI: 10.1016/0009-2614(69)80105-3.

[120] Martin Goez. "Elucidating Organic Reaction Mechanisms Using Photo-CIDNP Spectroscopy". In: *Hyperpolarization Methods in NMR Spectroscopy*. Ed. by Lars T. Kuhn. Vol. 338. Topics in Current Chemistry. Berlin, Heidelberg: Springer, 2013, pp. 1–32. DOI: 10.1007/128_2012_348.

[121] R. Kaptein. "Simple Rules for Chemically Induced Dynamic Nuclear Polarization". In: *Journal of the Chemical Society D: Chemical Communications* (14 1971), pp. 732–733. DOI: 10.1039/C29710000732.

[122] Roland Riek, Jocelyne Fiaux, Eric B. Bertelsen, Arthur L. Horwich, and Kurt Wüthrich. "Solution NMR Techniques for Large Molecular and Supramolecular Structures". In: *Journal of the American Chemical Society* 124.41 (2002), pp. 12144–12153. DOI: 10.1021/ja026763z.

[123] Richard R. Ernst. "Numerical Hilbert transform and automatic phase correction in magnetic resonance spectroscopy". In: *Journal of Magnetic Resonance* 1.1 (1969), pp. 7–26. ISSN: 0022-2364. DOI: 10.1016/0022-2364(69)90003-1.

[124] C Ammann, P Meier, and AE Merbach. "A simple multinuclear NMR thermometer". In: *Journal of Magnetic Resonance* 46.2 (1982), pp. 319–321. DOI: 10.1016/0022-2364(82)90147-0.

[125] Marco Tonelli et al. "Laser- and cryogenic probe-assisted NMR enables hypersensitive analysis of biomolecules at submicromolar concentration". In: *Proceedings of the National Academy of Sciences* 116.12 (2019), pp. 5556–5561. DOI: 10.1073/pnas.1819413116.

[126] Jonathan J. Helmus and Christopher P. Jaroniec. "Nmrglue: an open source Python package for the analysis of multidimensional NMR data". In: *Journal of Biomolecular NMR* 55.4 (2013), pp. 355–367. DOI: 10.1007/s10858-013-9718-x.

[127] H. W. Kuhn. "The Hungarian Method for the Assignment Problem". In: *Naval Research Logistics Quarterly* 2.1–2 (1955), pp. 83–97. DOI: 10.1002/nav.3800020109.

[128] Youngchun Kwon, Dongseon Lee, Youn-Suk Choi, Myeonginn Kang, and Seokho Kang. "Neural Message Passing for NMR Chemical Shift Prediction". In: *Journal of Chemical Information and Modeling* 60.4 (2020), pp. 2024–2030. DOI: 10.1021/acs.jcim.0c00195.

[129] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, and Pascal Friederich. "Graph Neural Networks for Materials Science and Chemistry". In: *Communications Materials* 3 (2022), p. 93. DOI: 10.1038/s43246-022-00315-6.

[130] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. "Neural Message Passing for Quantum Chemistry". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 1263–1272. URL: https://proceedings.mlr.press/v70/gilmer17a.html.

[131] Benson Chen, Regina Barzilay, and Tommi S. Jaakkola. *Path-Augmented Graph Transformer Network*. 2019. arXiv: 1905.12712 [cs.LG]. URL: https://arxiv.org/abs/1905.12712.

[132] NMRShiftDB Team. *NMRShiftDB2: Open Access NMR Database*. Accessed: 2025-02-21. 2025. URL: https://nmrshiftdb.nmr.uni-koeln.de/.

[133] Stefan Kuhn and Norbert E. Schlörer. "Facilitating quality control for spectra assignments of small organic molecules: nmrshiftdb2 – a free in-house NMR database with integrated LIMS for academic service laboratories". In: *Magnetic Resonance in Chemistry* 53.7 (2015), pp. 582–589. DOI: 10.1002/mrc.4263.

[134] Greg Landrum. *RDKit: Open-source cheminformatics software*. https://www.rdkit.org. 2023.

[135] Roberto Buratto, Daniele Mammoli, Elisabetta Chiarparin, Glyn Williams, and Geoffrey Bodenhausen. "Exploring Weak Ligand–Protein Interactions by Long-Lived NMR States: Improved Contrast in Fragment-Based Drug Screening". In: *Angewandte Chemie International Edition* 53.42 (2014), pp. 11376–11380. DOI: 10.1002/anie.201404921.

[136] Eric Jonas and Stefan Kuhn. "Rapid prediction of NMR spectral properties with quantified uncertainty". In: *Journal of Cheminformatics* 11.1 (2019), p. 50. DOI: 10.1186/s13321-019-0374-3.

[137] Mohammad Meshkian, Nicolas Schmid, Andreas Henrici, and Stefan Bruderer. "Analysis of 1D NMR Spectra with 2D Image Processing Techniques". In: *Physica Scripta* 100.2 (2025), p. 026011. DOI: 10.1088/1402-4896/ada595.

[138] Nicolas Schmid, Marc Wanner, Giulia Fischetti, Andreas Henrici, Mohammad Meshkian, Stefan Bruderer, Rudolf M. Füchslin, Bastian Heitmann, Jan D. Wegner, and Roland K. O. Sigel. "MolDeTr: A Chemistry-Informed Deep Learning Model for Next-Generation Automated Analysis of 1H NMR Spectra". Preprint, ChemRxiv. 2026. DOI: 10.26434/chemrxiv.10001683.v2.

[139] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*. Curran Associates, Inc., 2017, pp. 5998–6008.

[140] Xianfeng Zhai and Yaxiong Liu. "UncerTrans: Uncertainty-Aware Temporal Transformer for Early Action Prediction". In: *Scientific Reports* 16 (2026), p. 7068. DOI: 10.1038/s41598-026-38107-4.

[141] Oldřich Vyšata, David Matyáš, Rane Kinnari Chandrashekhar, and Aleš Procházka. "Comparative Analysis of CNN, LSTM, and Transformer Models for Gait Classification Using Single C7-Positioned IMU in Neurological Disorders". In: *Proceedings of the 25th International Conference on Digital Signal Processing (DSP)*. IEEE, 2025, pp. 1–5. DOI: 10.1109/DSP65409.2025.11074893.

[142] Jun-Jie Wang, Yongqi Jin, Chen-Yu Zhi, Yu-Jie Liu, Xu-Hao Huang, Fanjie Xu, Xiaohong Ji, Xi Fang, Haoyi Tao, Weinan E, Linfeng Zhang, Guolin Ke, and Rong Zhu. "NMRexp: A Database of 3.3 Million Experimental NMR Spectra". In: *Scientific Data* 12 (2025), p. 1954. DOI: 10.1038/s41597-025-06245-5.

[143] Luca G. Mureddu, Timothy J. Ragan, Edward J. Brooksbank, and Geerten W. Vuister. "Fragment-Based Drug Discovery by NMR: Where Are the Success Stories?" In: *Frontiers in Molecular Biosciences* 9 (2022), p. 834453. DOI: 10.3389/fmolb.2022.834453.

[144] Bruker BioSpin. *Fragment-Based Screening*. Accessed: 2026-03-09. 2026. URL: https://www.bruker.com/en/products-and-solutions/mr/nmr-software/fragment-based-screening.html.

[145] Chen Peng, Alexandra Frommlet, Sabine Rauscher, Martin Oesterle, Bernhard Faller, Michael Sattler, and Paul Schanda. "Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed NMR Binding Data". In: *Journal of Medicinal Chemistry* 59.7 (2016), pp. 3303–3310. DOI: 10.1021/acs.jmedchem.6b00019.

[146] Luca G. Mureddu, Timothy J. Ragan, Edward J. Brooksbank, and Geerten W. Vuister. "CcpNmr AnalysisScreen, a New Software Programme with Dedicated Automated Analysis Tools for Fragment-Based Drug Discovery by NMR". In: *Journal of Biomolecular NMR* 74.10–11 (2020), pp. 565–577. DOI: 10.1007/s10858-020-00321-1.

[147] C. S. Damberg, Vladislav Y. Orekhov, and Martin Billeter. "Automated Analysis of Large Sets of Heteronuclear Correlation Spectra in NMR-Based Drug Discovery". In: *Journal of Medicinal Chemistry* 45.26 (2002), pp. 5649–5654. DOI: 10.1021/jm020866a.

[148] Vincenzo Laveglia, Andrea Giachetti, Linda Cerofolini, Kevin Haubrich, Marco Fragai, Alessio Ciulli, and Antonio Rosato. "Automated Determination of Nuclear Magnetic Resonance Chemical Shift Perturbations in Ligand Screening Experiments: The PICASSO Web Server". In: *Journal of Chemical Information and Modeling* 61.12 (2021), pp. 5726–5733. DOI: 10.1021/acs.jcim.1c00871.

[149] Simon H. Rüdisser, Nils Goldberg, Marc-Olivier Ebert, Helena Kovacs, and Alvar D. Gossert. "Efficient affinity ranking of fluorinated ligands by ^{19}F NMR: CSAR and FastCSAR". In: *Journal of Biomolecular NMR* 74.10-11 (2020), pp. 579–594. DOI: 10.1007/s10858-020-00325-x.