Progress in Nuclear Magnetic Resonance Spectroscopy 148–149 (2025) 101575

Contents lists available at ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

journal homepage: www.elsevier.com/locate/pnmrs

Machine learning in NMR spectroscopy

Piotr Klukowski [a,*], Roland Riek [a,*], Peter Güntert [a,b,c,*]

[a] Institute of Molecular Physical Science, ETH Zurich, Zurich, Switzerland
[b] Institute of Biophysical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany
[c] Department of Chemistry, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

ARTICLE INFO

Editor: David Neuhaus

Keywords:
NMR spectroscopy
Machine learning
Deep learning
Automated spectrum analysis
Peak picking
Chemical shift assignment
Structure determination
Chemical shift prediction
Non-uniform sampling

ABSTRACT

NMR spectroscopy is a versatile technique for studies of molecular structures, dynamic processes, and intermolecular interactions across a broad range of systems, including small molecules, macromolecules, biomolecular assemblies, and materials in both solution and solid-state environments. As the complexity of NMR studies continues to pose challenges for practitioners, the integration of machine learning is recognized as a promising research direction for improving data acquisition, processing, and analysis. Here, we summarize recent findings in this area, highlighting common applications such as signal detection, chemical shift assignment, structure determination, chemical shift prediction, non-uniform sampling reconstruction, and denoising. For each of these applications, we discuss machine learning methods, design choices, and key publicly available data repositories. We conclude by identifying major trends and emerging directions at the intersection of machine learning and NMR spectroscopy that could help advance research in the field.

Contents

1. Introduction . . . 1
2. Automated analysis of experimental data . . . 3
   2.1. Peak picking . . . 3
   2.2. Protein chemical shift assignment . . . 4
   2.3. Protein structure determination . . . 5
3. Predicting NMR probe – Structure/dynamics-property relationships . . . 6
   3.1. Chemical shifts in proteins . . . 6
   3.2. Chemical shifts in small molecules and molecular solids . . . 7
   3.3. Scalar couplings in small molecules . . . 8
   3.4. Protein dynamics . . . 8
4. Spectrum reconstruction and quality enhancement . . . 9
   4.1. NUS reconstructions . . . 9
   4.2. Virtual decoupling . . . 10
   4.3. Denoising . . . 11
5. Conclusions and outlook . . . 11
   5.1. ML adoption in laboratory practice . . . 11
   5.2. Reducing experimental input . . . 12
   5.3. NMR experiment design optimization for deep learning analysis . . . 12
   5.4. Spectra databases . . . 12
   5.5. Toward comprehensive structure and dynamic determination . . . 12
   Glossary of abbreviations . . . 12
   CRediT authorship contribution statement . . . 12

* Corresponding authors at: Institute of Molecular Physical Science, ETH Zurich, Zurich, Switzerland.
E-mail addresses: piotr.klukowski@phys.chem.ethz.ch (P. Klukowski), roland.riek@phys.chem.ethz.ch (R. Riek), peter.guentert@phys.chem.ethz.ch (P. Güntert).

https://doi.org/10.1016/j.pnmrs.2025.101575
Received 14 May 2025; Accepted 25 June 2025
Available online 1 July 2025

Declaration of competing interest . . . 12
Acknowledgements . . . 12
Data availability . . . 12
References . . . 12

1. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy offers diverse opportunities for computational method development. This effort is driven by the continuous need to improve protocols for studying molecular systems at the atomic level. In recent years, the broader research landscape has been evolving rapidly, fueled by breakthroughs in artificial intelligence. These developments not only open new possibilities for molecular sciences but also highlight the pressing need for transformative improvements in long-standing NMR methodologies.

In the domain of data acquisition, advancements primarily focus on optimizing NMR measurement time. One important technique, Non-Uniform Sampling (NUS) [1], accelerates data acquisition by sampling a reduced number of datapoints in the indirect spectral dimensions while preserving information about the studied system as far as possible. The effectiveness of NUS, however, relies heavily on the reconstruction

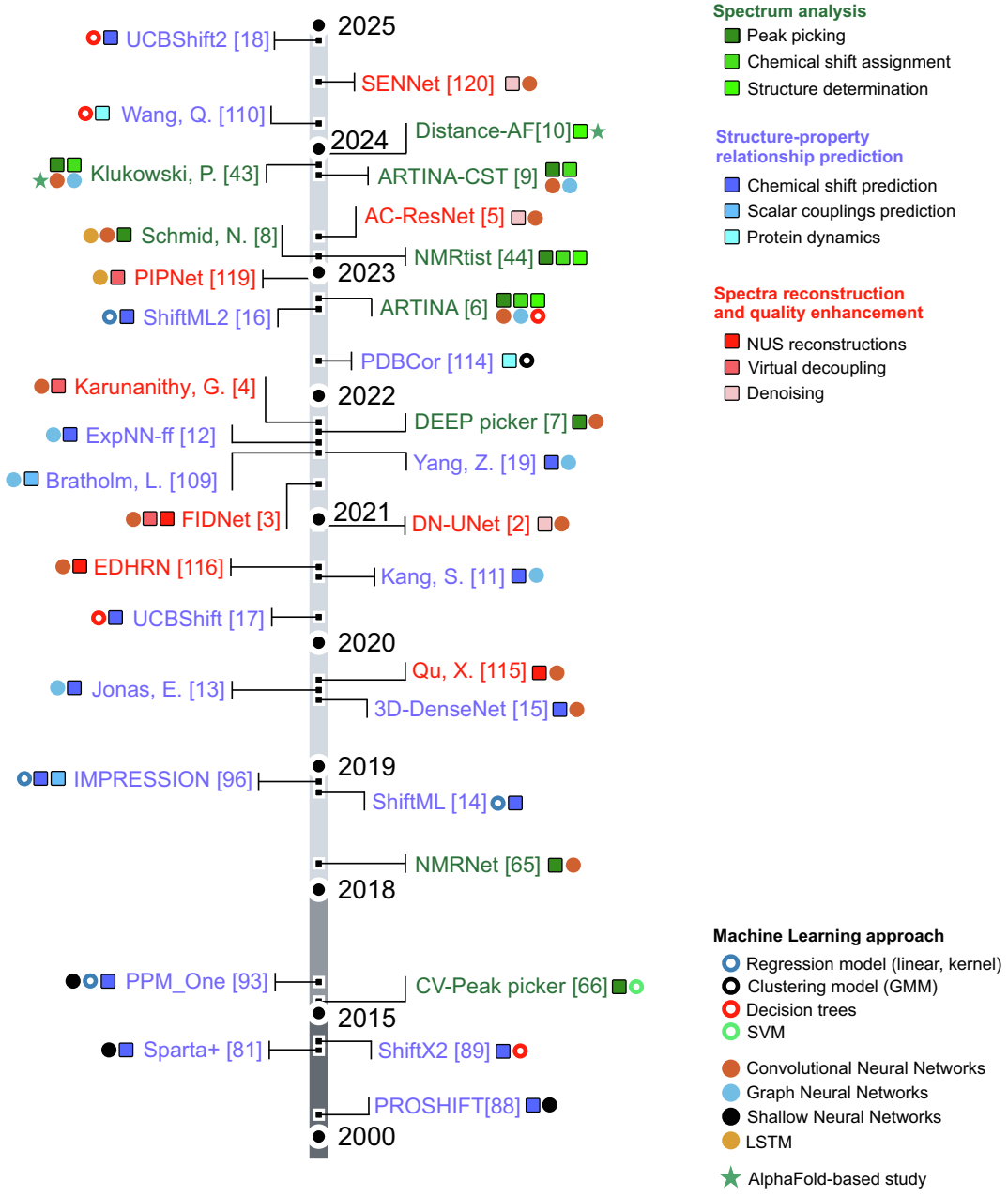

Fig. 1. Overview of articles highlighted in this literature review, arranged chronologically. Method names are color-coded according to their application areas. If a method name was not specified in the original work, the name of the first author is used instead. Symbols adjacent to each method name represent the primary modelling approach (Table 1) employed in the corresponding study.

Table 1
Common types of machine learning models applied in NMR spectroscopy.

	Underlying modelling approach	Typical application domains	NMR use cases
Gradient Boosted Trees (GBT)	Trains decision trees sequentially, where each subsequent tree is fitted to correct the residual errors of the previous ones. The prediction is made by combining the predictions of all trees in the sequence. The method presents an alternative for random forests.	Applied to classification and regression problems involving tabular datasets that contain numerical and categorical features relevant to the prediction task. Effective also for medium-sized datasets, provided that hyperparameters (such as the number of trees and tree depth) are appropriately tuned.	Used in ARTINA to evaluate protein structure proposals based on numerical features derived from peak picking, chemical shift assignment, and structure calculation, such as the number of long-range NOEs or the value of the CYANA target function.
Convolutional Neural Networks (CNN)	Applies learnable filters over input data arranged on a regular grid (e.g., images, spectral planes). Each filter is designed to detect local patterns such as peaks or edges by convolving across the data. The network consists of multiple layers of convolution and pooling operations, which progressively extract higher-level features.	Effective for analyzing grid-like data (e.g., 1D signals, 2D images, 3D volumes), where patterns are spatially localized but may appear at different positions. Widely used with images, videos, point clouds, spectra, and other data types with similar spatial or temporal structure.	Used for automated peak detection in NMR spectra by learning characteristic peak shapes and distinguishing them from artifacts, noise and baseline.
Message-passing Graph Neural Network (GNN)	Operates on graph-structured data, where nodes represent entities (e.g., atoms) and edges represent relationships (e.g., bonds, spatial proximity). Each node begins with a feature vector, and during each message-passing step, nodes aggregate information from their neighbors and update their own state accordingly. This process is repeated, allowing the model to learn representations of the entire graph or its substructures.	Designed for data where relationships between elements are best represented as a graph rather than a grid or sequence. Applicable to modelling biological networks, molecular structures in chemistry (e.g., for property prediction), and physical system simulations.	Extensively used in chemical shift prediction for small molecules. Given the input compound, represented as the molecular graphs, the model learns how atomic environments and molecule geometry relate to observed NMR chemical shifts.
Long Short-Term Memory networks (LSTM)	LSTM is a type of neural network that is designed to model sequential data. It maintains a memory cell that is updated through gated operations, allowing the model to preserve important information across long sequences while discarding irrelevant details.	Commonly used in tasks involving ordered or time-dependent data sequences, such as natural language processing, speech recognition, time series forecasting, and genomic sequence analysis. Particularly useful when the prediction depends on patterns spread over long distances in the input sequence.	Applied to the analysis of 1D NMR spectra, where it captures global relationships between signals along the spectral axis. It enables the model to interpret NMR signals based on wide spectral context, such as relative signal positioning.
Kernel regression	Makes predictions based on the similarity between data points, rather than learning explicit parameters. The model computes a weighted average of training examples, where the weights are determined by a kernel function that quantifies similarity between the input and each training point.	Regression tasks where feature vectors describing the input can be compared using a similarity measure. Commonly used when data is limited but reliable descriptors (e.g., distances, angles, atomic environments) are available. Applicable to regression problems where the underlying functional relationship is unknown or difficult to model explicitly.	Applied to chemical shift prediction in molecular solids, where atomic environments are encoded as feature vectors and shifts are predicted by averaging training values weighted by structural similarity.

methods used to recover missing data points, making it a key challenge for algorithmic and machine learning (ML) developments. Beyond spectral reconstruction strategies, the resolution and sensitivity of NMR spectra are fundamental properties that drive further research efforts. Various approaches have been explored to enhance these aspects, including denoising techniques [2], virtual decoupling [3,4], and information synthesis from sequential measurements [5], among others.

Another group of ML applications in NMR spectroscopy cluster around experimental data analysis. Traditionally, this process has relied on manual work or semi-automated methods, requiring expert knowledge and significant time investment. The complexity of data analysis varies depending on the type of investigation and the system under study. For small molecules, the diversity of chemical space can lead to ambiguities in spectral interpretation. In protein NMR, high signal overlap and the presence of thousands of recorded signals make the process labor-intensive, often creating bottlenecks in research workflows. In solid-state NMR, dipolar couplings, chemical shift anisotropy, and relaxation effects present additional challenges for the spectral analysis. As a result, the demand for advanced algorithms and machine learning models to automate NMR data analysis remains high. Focal areas include visual spectrum analysis [6–8], chemical shift assignment [6,9], and structure determination [6,10].

An important factor in NMR spectroscopy is the indirect nature of the observables in the measured data. Cross-peaks and chemical shifts recorded in NMR spectra must be accurately linked to molecular structure and properties to address research questions. Establishing a more direct link between a structure and NMR observables constitutes one of the largest application domains for ML methods. Dedicated techniques have been developed for both liquid- and solid-state measurements, covering wide classes of systems, including small molecules [11–13], molecular solids [14–16] and macromolecules [17–19], offering approaches for structure-based prediction of chemical shifts, coupling constants, and molecular dynamics effects.

Among the systems studied with NMR, proteins represent one of the key application domains. About 83% of all macromolecular structures deposited in the Protein Data Bank (PDB) database have been determined with X-ray and only about 6.3% by NMR (decaying from 10% a few years ago), despite the finding that about 80% of proteins and protein complexes apparently do not crystallize [20]. The methods for structural studies have been improved significantly in the last two decades, including in particular automation in X-ray crystallography at large-scale synchrotron facilities along with well-developed software tools [21] and the resolution revolution in cryo-EM (electron microscopy) with breakthroughs both in instrumentation and software making structure determination of large proteins (> 100 kDa) rather straightforward [22], albeit often at intermediate resolution [23]. In addition, the recent breakthrough in structure prediction, in particular by AlphaFold [24,25], has provided a huge number of structural models that "are valuable hypotheses and accelerate but do not replace experimental structure determination" [26]. In contrast, until recently—specifically, before the introduction of deep learning, a class of machine learning methods based on multi-layered (or "deep") neural network architectures that automatically learn patterns from large datasets—solution state NMR structure determination was still following the same principles and methods as 25 years ago [27–30]. Hence, it takes usually several months of manual work by a dedicated NMR spectroscopist to analyze the NMR spectra of a small to medium size protein (i.e., 3–30 kDa, eventually up to 50–70 kDa with appropriate approaches) and to obtain its three-dimensional structure, which is considerably longer than the measurement time. Likewise, the

Table 2
Highlighted ML approaches for the analysis of experimental NMR data.

Approach	Scope of application	Approach
ARTINA [6]	Peak picking with an emphasis on protein data Automated chemical shift assignment Chemical shift transfer Structure-based chemical shift assignment Protein structure determination with experimental data	Convolutional neural networks Graph neural networks Gradient boosted trees Metaheuristic search (genetic algorithm, simulated annealing)
DEEP picker [7]	Peak picking with an emphasis on metabolomics	Convolutional neural networks
Schmidt et al. [8]	Peak picking with an emphasis on 1D experiments and molecules	Convolutional neural networks, LSTM
Distance-AF [10]	Integrating experimentally observed restraintsw ith AlphaFold predictions	AlphaFold2 adaptation
Klukowski et al. [43]	Integrating automated NMR spectra analysisw ith AlphaFold2 predictions to reduce amount of experimental data required for automated chemical shift assignment	ARTINA integration with AlphaFold2 and UCBShift
Wetton et al. [9]	Chemical shift transfer	ARTINA adaptation For chemical shift transfer
NMRtist [44]	Online system that hosts ARTINA (nmrtist.org)	–

elucidation of protein dynamics using a plethora of methods [31–35], as well as highly sensitive probes such as Residual Dipolar Couplings (RDCs) [36–38] and exact Nuclear Overhauser Enhancements (eNOEs) [39,40], are time-intensive both at the NMR spectrometer and in the analysis.

The ability of neural networks to learn complex spectral patterns, infer missing information, and generalize across large datasets has led to progress in NMR spectroscopy. In this review, we highlight a selection of machine learning applications in NMR (Fig. 1), which employ common types of machine learning models (Table 1). We summarize the main developments in machine learning for experimental data analysis (chapter 2), in establishing relationships between NMR probes and structure/dynamics of the biomolecule under study (chapter 3), and in enhancing the information content of NMR experiments and accelerating their acquisition (chapter 4). After reviewing published methods within this context, we discuss open opportunities in the area and possible future research directions.

2. Automated analysis of experimental data

Conventional NMR spectrum analysis involves signal detection (peak picking), assigning signals to atoms in the studied system (chemical shift assignment), and subsequently using this information for structure, interaction and dynamics determination. Despite the large volume of NMR data that is acquired to address research problems in chemistry, structural biology, and drug discovery—with over 14,000 NMR structures of macromolecules deposited in the Protein Data Bank [41] and tens of thousands of NMR spectra of organic compounds uploaded to NMRShiftDB [42]—the data analysis workflow largely relies on manual work. Machine learning methods hold the promise of automating these tasks, reducing the time, resources and expertise required to analyze NMR spectra. In this chapter, we highlight a selection of ML approaches for the analysis of experimental NMR data (Table 2).

2.1. Peak picking

The first approaches to cross-peak detection emerged in the 1980s, relying on pattern recognition and automated extraction of geometrical features from NMR spectra [45–47]. Over time, various methods have been developed, with many clustering around matrix decomposition techniques (AUTOPSY [48], three-way decomposition [49,50], singular value decomposition [51], non-negative matrix factorization [52]), thresholding (NMRView [53,54], XEASY [55], CCPN [56]), Bayesian approaches [57–59], and geometrical feature extraction [60–62]. Alongside non-ML solutions, machine learning approaches to cross-peak detection have evolved, initially taking the form of shallow, fully-connected neural networks [45,63,64] before advancing into more sophisticated models that are now among the leading methods in the field [6–8,65,66].

NMRNet [65] formulates cross-peak detection in NMR spectra as a binary classification task. Initially, the method analyzes the spectrum to identify key points that could represent recorded NMR signals. Subsequently, 2D image patches of size 48 × 48 points are cropped around each key point and used as input for prediction with a Convolutional Neural Network (CNN, Table 1). The model predicts the likelihood that the input image contains a true peak rather than an artifact. The architecture was trained and evaluated exclusively on experimental data that consists of 71 2D and 3D NMR spectra collected on 14 proteins.

Another application of convolutional neural networks in NMR spectroscopy is DEEP picker [7]. Unlike NMRNet, which formulates peak detection as a binary classification task based on localized 2D image patches, DEEP picker employs a sequence of 1D convolutional layers to extract spectral features along each spectrum dimension. The model, consisting of eight hidden layers, processes 1D spectral cross-sections independently before integrating predictions to reconstruct signals in higher-dimensional spectra. A key advantage of DEEP picker is its ability to predict not only peak positions but also signal amplitudes and shapes, providing a richer characterization of spectral features. While the approach is adaptable to spectra of arbitrary dimensionality, the authors focused on its application to 2D spectra, with an emphasis on protein and metabolomics data. The model was trained on simulated 1D spectra and validated on an experimental test set, demonstrating generalization to real measurements, such as [^{13}C,^{1}H]-HSQC, [^{15}N,^{1}H]-HSQC, ^{1}H–^{1}*H* NOESY and ^{1}H–^{1}H TOCSY experiments.

One of the latest deep learning models for cross-peak detection is integrated into the ARTINA method [6]. This approach leverages two deep residual neural networks (a type of CNN) to perform peak picking in 2D–4D NMR spectra. The first network performs binary classification, analyzing 2D spectral patches cropped from the input spectrum to detect the approximate positions of peaks by evaluating the proximity of signal extrema. The second network refines these initial detections through a regression-based approach, deconvolving and precisely localizing peaks initially identified as true signals by the first model. The visual spectrum analysis module within ARTINA was developed and tested on 1329 experimentally acquired 2D–4D NMR spectra, demonstrating its ability to automate peak detection across a broad range of protein NMR datasets. The method operates fully autonomously, requiring only the spectrum as input, and is publicly accessible as a web service within the NMRtist platform [44] (nmrtist.org). Additionally, the dataset used to develop ARTINA is publicly available [67] (nmrdb.ethz.ch), making it possible to reproduce 100 protein structures and assignments directly from NMR spectra. This dataset consists of experimentally acquired 2D–4D NMR spectra, manually solved protein structures (typically deposited in the PDB), manually assigned chemical shifts (from BMRB), lists of distance restraints, and expected cross-peak coordinates. The resource is designed to serve for the development of deep learning

methods in NMR spectroscopy, benchmarking NMR software, and educational purposes.

While previously discussed methods focus on peak detection in multidimensional protein NMR spectra, deep learning has also been applied to the analysis of one-dimensional experiments, particularly for small molecules. Schmidt et al. [8] proposed an approach that leverages a dataset of 250,000 simulated spectra, generated using a custom simulation framework developed in their study. The model architecture integrates an inception layer (a commonly used building block of CNNs) for initial feature extraction from normalized spectral data, followed by two fully connected layers to enhance representation learning. This is further refined by a Long Short-Term Memory (LSTM, Table 1) network that captures global correlations across the spectrum, and a prediction head (the final part of the neural network that computes the output values) comprising three fully connected layers. The model is trained to predict peak positions and corresponding signal widths. Performance evaluation was conducted on both simulated and experimental datasets, with the latter including 10 ^{1}H 1D spectra recorded on 400–600 MHz instruments, one ^{13}C 1D spectrum, and one benchtop spectrum. The performance of the method was evaluated using custom metrics designed by the authors to assess peak detection quality: (i) the fraction of annotated peaks (ground truth) that is picked correctly, (ii) a sparsity penalty, derived from a ratio between the number of annotated and predicted peaks, and (iii) a spectrum reconstruction error. On simulated test data, the method achieved scores (i–iii) of 0.92, 1.00 and 0.98, respectively, while on experimental data, these scores were 1.00, 0.88, and 0.97. These results indicate only a modest performance drop between simulated and real-world data, highlighting the model's robustness and generalization capabilities.

2.2. Protein chemical shift assignment

Chemical shift assignment has been a long-standing challenge in NMR spectroscopy, traditionally addressed by algorithms that rely on manually curated peak lists and heuristics. Among the widely used methods, FLYA [68], PINE-NMR [69], AUTOASSIGN [70] and MARS [71] employ network-based optimizations, probabilistic frameworks, or rule-based decision-making to match observed NMR resonances with atomic nuclei in the protein structure. While these approaches have significantly reduced the manual effort required for assignment, they often struggle with ambiguous signals, incomplete data, and variations in peak list quality. To overcome these limitations, recent advancements have introduced deep learning-based approaches that leverage learned priors from large datasets to guide the assignment process and improve robustness.

In the ARTINA method [6], the authors combine FLYA with a Graph Neural Network (GNN, Table 1). This hybrid approach takes as input either manually curated peak lists or automatically detected cross-peaks, which are identified by deep convolutional neural networks as discussed in the previous section. First, the FLYA algorithm generates an initial chemical shift assignment based on provided peak coordinates. Then, the GNN component enhances this assignment by incorporating contextual information from thousands of previously assigned proteins archived in the BMRB database, allowing it to predict expected values for so far unassigned or statistically weakly assigned atoms, which are used to augment the input for running FLYA a second time. This iterable refinement process results in improved assignment accuracy. The approach has been validated using a dataset of experimentally acquired NMR spectra [67], allowing the automated reproduction of chemical shift assignments for 100 proteins from the original measurements and their comparison with manual assignments. The method achieved an overall median chemical shift assignment accuracy of 91.4%, where accuracy was defined as the percentage of chemical shift assignments that match the manually assigned chemical shifts within specified tolerances (^{1}H: 0.035 ppm, ^{13}C/^{15}N: 0.4 ppm), accounting for chemically equivalent atoms and symmetry-related ambiguities. The most significant challenges were observed for long, charged side chains and aromatic rings, where assignment ambiguity tends to be higher.

Frequently, the chemical shift assignment of a homologous protein or the same protein studied under different experimental conditions is already available. This prior knowledge enables the use of chemical shift transfer, where the assignment of a related protein serves as a starting point for the target protein, thereby eliminating the need for *de novo* assignment. Chemical shift transfer is particularly beneficial in studies of proteins in various buffer environments or for protein interactions, where binding events or conformational changes can induce localized chemical shift perturbations, making it essential to track variations while leveraging existing assignments. A follow-up study from the ARTINA project expanded the scope of chemical shift transfer [9], demonstrating that even limited experimental datasets—such as [^{13}C,^{1}H]-HSQC, [^{15}N,^{1}H]-HSQC, ^{13}C-resolved ^{1}H–^{1}H NOESY, and ^{15}N-resolved ^{1}H–^{1}H NOESY—can effectively facilitate the transfer of both backbone and side-chain chemical shift assignments. The method was validated through multiple numerical experiments, including simulated perturbations of chemical shifts and analyses of homologous protein pairs extracted from the BMRB database. Across all experiments, chemical shift transfer consistently outperformed *de novo* assignment, with its accuracy improvement proportional to the magnitude of chemical shift perturbation between the source and target proteins. A key advantage of this method is its ability to directly utilize NMR spectra of the target system in conjunction with the chemical shift list of the source protein for transfer, while still allowing for the use of manually curated peak lists when necessary.

Deep learning-based protein structure prediction serves as a valuable source of prior information, significantly reducing ambiguity in the analysis of experimental NMR data. By leveraging these predictions, spectroscopists can reach biologically relevant conclusions with fewer NMR spectra and shorter measurement times. For *de novo* chemical shift assignment with ARTINA using only spectra and sequence as input, an average of 13.3 spectra, predominantly 3D experiments, were used to automatically reconstruct protein assignments. To systematically assess the impact of *in silico* protein predictions on chemical shift assignment accuracy and data reduction, a large-scale study with over 5000 automated protein assignments was conducted [43]. The core strategy involved utilizing subsets of experimentally acquired NMR spectra in conjunction with prior information from AlphaFold2-predicted structural models and UCBShift-derived [17] chemical shift predictions. Notably, the study demonstrated that using just five 3D spectra, complemented by AlphaFold and UCBShift predictions, achieved higher automated assignment accuracy (92.6%) than utilizing the full spectral dataset without these predictive tools (91.4%).

2.3. Protein structure determination

Machine learning plays an increasingly important role in protein NMR structure determination, with its applications spanning across two primary areas, which balance differently the reliance on experimental and *in silico* predictions: (a) structure determination based exclusively on the experimental data, and (b) using sparse experimental data in the context of *in silico* structure elucidation.

The ARTINA method exemplifies the first application area, where the sole role of machine learning is to automatically analyze the experimental data, simplifying and accelerating the process. ARTINA [6] is composed of three major logical blocks: spectrum analysis using deep neural networks (CNNs, as discussed in the peak picking section above), a chemical shift refinement cycle (detailed in the chemical shift assignment section above), and a structure calculation cycle. During the structure determination cycle, ARTINA utilizes chemical shift lists and NOESY peak lists to generate ten structural proposals. These proposals are ranked using Gradient Boosted Trees (Table 1)—a machine learning method that combines the predictions of many decision trees, trained sequentially to minimize the prediction error. The best-performing

Table 3
Highlighted ML approaches for structure-based chemical shift prediction in proteins, molecular solids and small molecules.

Method	Model / approach	Primary dataset used	Prediction scope
PROSHIFT [88] (2003)	Neural network Feed forward neural network with single hidden layer that predicts chemical shift values based on 350 handcrafted features	322 protein structures database with associated chemical shifts	Proteins, solution NMR ^{1}H, ^{13}C and ^{15}N
Sparta+ [81] (2010)	Neural network Feed forward neural network with single hidden layer that predicts chemical shift values based on 113 handcrafted features	580 protein structures database with associated chemical shifts, compiled based on SPARTA [84], TALOS+ [90] and Promega [91] studies.	Proteins, solution NMR H^{N}, H^{α}, C', C^{α}, C^{β}, N
ShiftX2 [89] (2011)	Regression trees, alignment Ensemble of regression trees combined with query sequence alignment against the database of chemical shifts.	322 protein structures database with associated chemical shifts, compiled based on RefDB [92], SPARTA [84] and SHIFTX [83].	Proteins, solution NMR ^{1}H, ^{13}C and ^{15}N
PPM_One [93] (2015)	Neural network, linear regression Feed forward neural network with single hidden layer (backbone chemical shift), linear regression (side-chain shifts)	405 protein structures database with associated chemical shifts	Proteins, solution NMR ^{1}H, ^{13}C and ^{15}N
SHIFTML [14] (2018)	Kernel regression models Gaussian process regression with SOAP kernel [94]	2500 structures selected from CSD [95] with DFT-calculated chemical shifts.	Molecular solids, ssNMR ^{1}H, ^{13}C, ^{15}N, ^{17}O
Jonas [13] (2019)	Deep neural network Graph convolutional neural network with input represented as molecular graph with 37 atom features.	32,538 small molecules ($\leq$ 64 atoms) selected from NMRShiftDB2 [42]	Small molecules Solution NMR ^{1}H, ^{13}C
3D-DenseNet [15] (2019)	Deep neural network 3D-DenseNet with input represented as 3D grid that presents atom neighborhood at different spatial resolutions	2500 structures selected from CSD [95] with DFT-calculated chemical shifts, compiled based on ShiftML	NMR crystallography ^{1}H, ^{13}C, ^{15}N, ^{17}O
UCBShift [17] (2020)	Random forests, alignment Ensemble of regression trees combined with query sequence alignment against the database of chemical shifts.	847 protein structures database with associated chemical shifts, compiled based on SPARTA+, ShiftX2 and RefDB.	Proteins, solution NMR H^{N}, H^{α}, C', C^{α}, C^{β}, N
Kang [11] (2020)	Deep neural network Message passing graph neural network with input represented as molecular graph (node and edge features), trained with weakly supervised learning.	32,538 small molecules ($\leq$ 64 atoms) selected from NMRShiftDB2 [42], compiled based on Jonas et al. [13]	Small molecules Solution NMR ^{1}H, ^{13}C
Impression [96] (2020)	Kernel regression models Kernel Ridge Regression with Coulomb matrix [97], aSLATM [98] and FCHL [99] kernel similarity measures.	1292 structures from CSD [95] with DFT-calculated chemical shifts and J-couplings; 608 experimental $^{1}J_{CH}$ values from Venkata et al. [100]; 734 ^{1}H chemical shifts from Smith and Goodman [101]	Chemical structures comprising C, H, N, O, F atoms; solution-state NMR Coupling constant: $^{1}J_{CH}$, Chemical shifts: ^{13}C, ^{1}H
Yang et al. [19] (2021)	Deep neural network Message passing graph neural network trained in a generic framework that relies only on the geometry of the molecular graph	2664 protein structures with experimentally assigned chemical shifts: RefDB [92], ShiftX2 [89]; 369 metabolites from HMDB 4.0 [102].	Arbitrary chemical structures, solution NMR ^{1}H, ^{13}C,^{15}N
ExpNN-ff [12] (2021)	Deep neural network Message passing graph neural network, variation of Schnet architecture [103] with edge updates.	8000 structures from NMRShiftDB with DFT-calculated chemical shifts; 5000 structures with experimental ^{13}C chemical shifts; 24 structures from CHESHIRE dataset	Organic compounds, Solution NMR ^{1}H, ^{13}C
ShiftML2 [16] (2022)	Kernel regression models Ensemble of 32 kernel ridge regression models with SOAP kernels	16,013 structures (< 200 atoms in unit cell) from CSD [95] with DFT-simulated chemical shifts.	Molecular solids, ssNMR ^{13}C, ^{15}N, ^{17}O, ^{19}F, ^{33}S, ^{31}P, ^{35}Cl, ^{23}Na, ^{43}Ca, ^{25}Mg, ^{39}K
UCBShift 2.0 [18] (2024)	Random forests, alignment Random forest regression combined with query sequence and structure alignment against the database of chemical shifts.	1051 protein structures with associated experimental chemical shits, compiled based on SPARTA+, ShiftX2 and RefDB.	Proteins, Solution NMR ^{1}H, ^{13}C and ^{15}N

structure is either returned as the final output or fed back as contextual information to improve the subsequent chemical shift refinement cycle. The ARTINA structure calculation approach has been validated using a dataset of experimentally acquired NMR spectra [67], allowing the automated reproduction of 100 protein structures from original measurements. These structures were compared against corresponding Protein Data Bank (PDB) depositions, achieving a median root-mean-square deviation (RMSD) of 1.44 Å and mean RMSD of 1.55 Å. Importantly, 97% of the automatically solved structures exhibited an RMSD below 3 Å, with the remaining 3% primarily encountering challenges in accurately positioning the first or last secondary structure elements. Nonetheless, the overall protein folds closely resembled those obtained through manual structure determination [6].

The ARTINA method has been made publicly available through the NMRtist online platform [44], enabling fully automated peak picking, chemical shift assignment and structure calculation directly from protein NMR spectra. Its practical applications span studies of novel protein topologies [72], complex formation [73], pathogens [74–76], cancer research [77], immunoglobulins [78], allostery [79] and intracellular delivery [80].

Concerning the second area of application, outstanding developments in deep learning in silico protein structure predictions, such as AlphaFold2 [25], made it possible to design approaches in which the protein fold is primarily predicted in silico with additional information from sparse experimental data. An example of such an approach is Distance-AF [10]. The method utilizes the protein sequence and distance restraints, which may originate from NMR or other sources, as input. The process begins with generating an initial structure prediction using AlphaFold2, which serves as the starting point. Subsequently, the experimental distance restraints are linked to the predicted structure through a custom loss function that quantifies the discrepancy between the AlphaFold2 prediction and the provided restraints. This information is iteratively backpropagated to update the parameters of the Invariant Point Attention (IPA), a component of the AlphaFold2 structure module.

The iterative process continues until the predicted structure satisfies the specified distance restraints, effectively employing an overfitting mechanism that generates a unique set of parameters tailored to the target protein and its associated distance restraints. The custom loss function consists of four components: (a) quantification of discrepancies between the provided distance restraints and the corresponding distances in the predicted structure, (b) modification of AlphaFold2's Frame Aligned Point Error (FAPE) loss to preserve local structural integrity during iterative optimization, (c) restraints on backbone and sidechain torsion angles, and (d) penalties to prevent structural violations such as non-ideal bond lengths, bond angles, and atomic clashes. The performance of Distance-AF was evaluated *in silico* using 38 protein structures from the Protein Data Bank (PDB), including 25 monomeric proteins for which AlphaFold2 initially made predictions with root-mean-square deviation values exceeding 10 Å. These cases represent instances where AlphaFold2 predictions diverged substantially from the experimental structures, thus simulating conditions in which the restraints effectively override incorrect predictions. This setup enabled an evaluation of how Distance-AF handles conflicting information, and the results demonstrated that incorporation of distance restraints led in many cases to improvements in structural accuracy over AlphaFold2.

3. Predicting NMR probe – Structure/dynamics-property relationships

In NMR spectroscopy, the experimentally observed chemical shifts, coupling patterns, and related signal features provide indirect insights into a molecule's underlying structure and dynamics. Because these observables cannot be directly translated into structural or dynamic information, substantial effort, including the peak picking, shift assignment and additional analytical steps, is typically required to build reliable structural models or address biological questions. A more direct link between NMR observables and molecular properties could not only streamline the interpretation process—reducing reliance on extensive modelling and computational analyses—but also create new avenues for studies of molecular systems and facilitate rapid experimental validation of *in silico* predictions. Efforts toward this goal include developing solutions for structure-based chemical shift prediction, estimating J-couplings from atomic arrangements, and inferring dynamic information directly from NMR-derived structure models.

3.1. Chemical shifts in proteins

Chemical shifts are fundamental observables in NMR spectroscopy that encode the local physical environment of the studied system. Each chemical shift value is highly sensitive to the electronic surroundings of nuclei, reflecting variations in bonding, electronegativity, and the spatial arrangement of atoms. This sensitivity allows chemical shifts to serve as precise indicators of the chemical structure, secondary structure elements and to detect subtle conformational changes. Predicting chemical shifts from biomolecular structures has been a longstanding area of research explored both with machine learning [17,81,82] (Table 3) and algorithmic approaches [83–87]. A foundation for research on this topic was laid by methods such as PROSHIFT [88], Sparta+ [81] and ShiftX2 [89]. The first two approaches employ a fully connected neural network with a single hidden layer, coupled with hand-crafted features that capture the protein's local geometrical and chemical environment, to predict backbone chemical shifts from structural models. The last method, ShiftX2, uses a hybrid approach combining a chemical shift database search with predictions made using an ensemble of regression trees.

A recent method using structure-based chemical shift prediction, UCBShift 2.0 [18], employs an ensemble approach that integrates both alignment-based and structure-based predictions, revisiting the strategy previously explored in UCBShift 1.0 [17] and ShiftX2 [89]. The method consists of two complementary components: UCBShift-Y, which aligns the query protein to a database of manually assigned chemical shifts, and UCBShift-X, which utilizes random forest models trained on handcrafted features extracted from protein structures. This combined strategy enhances predictive accuracy by leveraging both homology-based information and machine learning-driven prediction. UCBShift 2.0 was developed using a dataset of 1051 manually assigned proteins—comprising 851 proteins for training and 200 for testing—which was curated from BMRB, Sparta+, RefDB, and ShiftX2 datasets. The method provides chemical shift predictions for ^{1}H, ^{13}C, and ^{15}N nuclei, covering both backbone and side-chain atoms.

Although deep learning is not yet dominating the field of structure-based chemical shift prediction of proteins, emerging approaches from this domain present several potential advantages over traditional machine learning techniques. One key benefit is the possibility to train generic models directly on the molecular graph geometry, which inherently encodes the structural and electronic environment of atoms, rather than relying on handcrafted features that limit the model application to a narrow class of systems compatible with the proposed feature extraction. The concept of training a generic model on the molecular graph geometry was explored by Yang et al. [19], who employed a message-passing graph neural network (GNN, Table 1) to predict ^{1}H, ^{13}C, and ^{15}N chemical shifts in proteins and beyond. The model uses as input a molecular graph representation, capturing both through-bond and through-space interactions. As this method does not require protein- or amino acid-specific descriptors, the authors were able to augment their training dataset—primarily derived from RefDB and ShiftX2—with additional metabolomics data [102]. The approach demonstrated performance comparable to state-of-the-art methods, such as UCBShift and SHIFTX+, while offering a more generalizable framework for chemical shift prediction. Notably, a phenomenological analysis of the GNN's learned representations revealed that the model inferred the sensitivity of chemical shifts to protein secondary structure, despite not being explicitly provided with secondary structure information, highlighting its ability to capture fundamental physicochemical relationships through learning alone.

3.2. Chemical shifts in small molecules and molecular solids

Unlike proteins, where chemical shifts are influenced by secondary structure and backbone interactions of amino acids, chemical shift prediction in small molecules and molecular solids requires machine learning models to capture diverse molecular scaffolds, functional groups, and local electronic environments. A contribution in this area was made by Jonas et al. [13] who introduced a Graph Convolutional Neural Network (GCNN) for predicting ^{1}H and ^{13}C chemical shifts in small molecules with up to 64 atoms. In this approach, molecules are represented as graphs, where each atom is described by 37 distinct features, such as hybridization state and formal charge, while chemical bonds are modeled as edges without additional attributes. The authors adapted the GCNN architecture proposed by Kipf and Welling [104] by modifying the layer-wise propagation equation to account for different types of chemical bonds within the molecular graph. Their final model consists of ten graph convolutional layers, followed by two prediction heads that estimate both the expected chemical shift values and their associated uncertainty. Two separate instances of the model were trained using experimental data from the NMRShiftDB2 database [42] to predict ^{1}H and ^{13}C chemical shifts independently. An important feature of the approach is uncertainty estimation, allowing the model to estimate the confidence of its predictions. The method achieved mean absolute errors (MAE) of 0.71–1.43 ppm for ^{13}C and 0.09–0.28 ppm for ^{1}H, where the lower limit applies to predictions made with the highest certainty (10% of the dataset) and the upper limit to all predicted chemical shifts. Additionally, the study demonstrated that the GCNN-based method outperforms Density Functional Theory (DFT) calculations across all types of chemical shifts.

ShiftML [14] addresses the problem of predicting chemical shifts in

molecular solids, a critical task in validating structural assignments, distinguishing known polymorphs, and determining crystal structures *de novo* from powders. The authors proposed a Gaussian Process Regression (GPR) model (kernel regression, Table 1) with Smooth Overlap of Atomic Positions (SOAP) multi-kernel, where each atomic environment is represented as a three-dimensional neighborhood density, constructed from a superposition of Gaussians centered at atomic positions within a defined cutoff radius. Trained and tested on 2500 examples extracted from the Cambridge Structural Database (CSD) with DFT-calculated chemical shifts, the model achieved root-mean-square errors of 0.49 ppm for ^{1}H, 4.3 ppm for ^{13}C, 13.3 ppm for ^{15}N, and 17.7 ppm for ^{17}O, relative to the DFT-calculated values. Additionally, the model demonstrated scalability to larger systems, with successful predictions for six structures containing 768 to 1584 atoms, highlighting its applicability to complex molecular solids. In evaluations against experimental data, ShiftML achieved accuracy comparable to DFT-based predictions (RMSE 0.39 vs. 0.38 ppm for ^{1}H), while being faster by several orders of magnitude.

The follow-up work, ShiftML2 [16], builds upon its predecessor in several important areas. First, the study covers a broader range of compounds, comprising up to 12 elements (i.e., H, C, N, O, S, F, P, Cl, Na, Ca, Mg, K), compared to 5 (i.e., C, H, N, O, S) in ShiftML. Secondly, ShiftML2 has the capacity to calculate chemical shifts for distorted structures, such as those not optimized using DFT, and snapshots obtained in MD simulations. The improved solution is an ensemble of 32 ridge regression models (kernel regression, Table 1), replacing the original Gaussian process regression, but preserving SOAP kernels for molecular representations. The prediction is made by averaging predictions of individual models, whereas prediction uncertainty is handled by scaled standard deviations of these predictions. The model was established with over 16,000 structures selected from CSD using farthest point sampling. It systematically outperforms its predecessor on benchmarked data, with minor improvements on the original ShiftML1 test set (0.48 ppm vs. 0.46 ppm proton RMSE), but major improvements in areas not covered by the previous study. For example, the authors used DFT to compute ^{1}H shieldings for up to 50 snapshots taken every 100 fs from MD simulations of 3 compounds. The average RMSE along MD trajectories was only slightly above the RMSE calculated for the relaxed structures, and substantially below the one from ShiftML1 (0.53 vs. 0.98 ppm proton RMSE).

Another approach, 3D-DenseNet [15], touched upon predicting chemical shifts in NMR crystallography. The authors utilized a dataset consisting of molecular crystal structures deposited in the Cambridge Structural Database [95] and DFT calculations of 2500 structures, compiled based on ShiftML1, to train the model. The molecular data is represented as a 3-dimensional grid centered at the atom that undergoes prediction, which captures the local chemical environment at different spatial resolutions. The deep learning architecture selected for the study, DenseNet [105], was developed in computer vision research, where it is a popular choice for image analysis and object recognition. DenseNet is a variant of convolutional neural networks that contains “dense blocks”, where each layer is connected to its preceding layers, such that for each layer, the feature-maps of preceding layers are used as inputs. This architectural design addresses the vanishing gradient problem, strengthens feature propagation, encourages feature reuse in the model inference and reduces the number of required model parameters. In the context of chemical shift prediction, the authors trained four instances of DenseNet to address predictions of ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shifts separately. By leveraging deep learning the authors demonstrated prediction accuracy that outperforms previously published classical machine learning approaches, such as kernel regression [14]. The 3D-DenseNet achieved RMSEs of 0.37, 3.3, 10.2 and 15.3 ppm for ^{1}H, ^{13}C, ^{15}N and ^{17}O respectively.

Another approach to chemical shift prediction has been demonstrated by Guan et al. [12]. The authors employed message passing graph neural networks to predict ^{1}H and ^{13}C chemical shifts in molecules. Their method design involved training the initial model (DFTNN) with DFT-optimized structure geometries and DFT-calculated chemical shifts. DFTNN already exhibits competitive performance, achieving RMSE values of 2.15 ppm for ^{13}C and 0.16 ppm for ^{1}H. To further refine the accuracy, the authors performed transfer learning with experimental data, focusing on ^{13}C only, to compensate for the inaccuracy of DFT-based chemical shifts in the initial training set. In the final step, additional transfer learning was executed to eliminate reliance on DFT-optimized geometries, enabling predictions using only molecular mechanics, which significantly reduces computational overhead. The source data for this study was obtained from NMRShiftDB by selecting 8000 structures with molecular weight up to 500 Da with the farthest-neighbor algorithm, and from the CHESHIRE [106] set of 24 organic molecules. On the test set, this approach achieves a notable MAE of 1.5 ppm for ^{13}C chemical shift predictions. The publication [12] demonstrates example applications of the proposed model, such as: (a) chemical structure determination by selection of the correct organic structure from candidates using experimental spectra, (b) automated detection and validation of data deposited in public NMR repositories, and (c) determination of chemical properties like sites for electrophilic aromatic substitution.

An interesting approach to chemical shift prediction was presented by Kang et al. [11]. The authors utilized weakly-supervised learning to predict chemical shifts in small molecules (up to 64 atoms). In weakly-supervised learning the model is trained to predict more detailed information than is available for all or some of the training examples. The most common approach in chemical shift prediction is to use supervised learning, an approach where the target chemical shifts are known for every atom of interest in the input molecule (e.g., manually assigned, DFT-calculated). Then, the ML model is trained to predict values that are directly compared with those available in the training set. In the proposed weakly-supervised approach, the training dataset consists of a list of shifts observed for a molecule (e.g., ppm values of peaks observed in a ^{1}H spectrum), but without their assignment to atoms. However, the model has the capacity to predict atom-level chemical shift values (i.e., with assignments). The authors presented how weakly-supervised techniques can increase the performance of supervised learning models (models can be trained with a mix of supervised and weakly-supervised training examples). This introduces a viable approach to incorporate more experimental data in ML modelling, as the amount of publicly available data suitable for weakly-supervised learning (signals with known shifts) is larger than that for supervised learning (the subset of signals with known shifts that have been assigned).

3.3. Scalar couplings in small molecules

J-couplings offer complementary structural information to chemical shifts in small molecules by encoding through-bond interactions that define bond connectivity, dihedral angles, and conformational preferences. The importance of accurate *J*-coupling prediction was highlighted in 2019, when a Kaggle competition [107] attracted 2700 data science teams from 84 countries to develop machine learning models for predicting $^{1}J_{HC}$, $^{1}J_{HN}$, $^{2}J_{HH}$, $^{2}J_{HC}$, $^{2}J_{HN}$, $^{3}J_{HH}$, $^{3}J_{HC}$ and $^{3}J_{HN}$ *J*-coupling constants in small molecules containing up to 9 heavy atoms (QM9 dataset [108]). The competition was based on DFT-simulated *J*-coupling constants and resulted in 47,800 machine learning model predictions submitted by participants. The findings were later summarized in an open-access manuscript [109]. The winning solution leveraged a graph-based representation, designed to explicitly capture atoms, bonds, atom triplets and others within a single data structure. The core model architecture was a graph transformer (a type of graph neural network). Following a common practice in Kaggle competitions, the authors built an ensemble of 13 variants of the architecture to further boost the performance.

IMPRESSION [96] is a method predicting ^{1}H, ^{13}C chemical shifts and $^{1}J_{CH}$ coupling in small molecules. The approach evaluated in the study is

kernel ridge regression (kernel regression, Table 1) with three molecular representations: Coulomb matrices, aSLATM and FCHL. It has been trained on a subset of 882 structures selected with active learning from the Cambridge Structural Database. The authors evaluated their approach on several experimental and DFT-calculated benchmarks, including 410 chemical structures harvested from CSD-500, 608 experimental $^{1}J_{CH}$ values from Venkata et al. [100] and 734 ^{1}H chemical shifts from Smith and Goodman [101]. On the test dataset, the reported MAE is 0.23 ppm/2.45 ppm/0.87 Hz for ^{1}H, ^{13}C chemical shifts and $^{1}J_{CH}$ coupling respectively, relative to the DFT-calculations.

3.4. Protein dynamics

Motion is an important aspect that shapes the role of proteins in biological processes. NMR spectroscopy, being able to conduct measurements in nearly physiological conditions, is a leading approach for acquiring protein dynamics data. Internal motion on the picosecond to nanosecond scale is mainly associated with vibration of chemical bonds, rotation of amino acid side chains, and loop motion. On the microsecond to millisecond timescale, the motion primarily involves conformational changes, movements between protein domains and folding/unfolding processes. Longer timescales still are primarily associated with processes like protein aggregation and large-scale conformational changes.

In a recent study, Wang et al. [110] proposed a machine learning solution that facilitates studies of fast dynamics (picosecond-nanosecond timescale). The idea is that fast protein motion is commonly studied by relaxation measurements using the model-free method [111] that quantifies the amplitude of local atomic motion through the order parameter S^2. However, experimental determination of S^2 is tedious as it requires measurements of (usually ^{15}N) longitudinal relaxation rates R_1, transverse relaxation rates R_2, and heteronuclear NOEs. Instead, in the proposed approach, the orders parameter of backbone ^{1}H–^{15}N bond directions are predicted using a random forest (50 trees) directly out of the NMR structure ensemble, without using any other experimental data. The model input consists of a set of handcrafted features extracted from the protein structure bundle, such as variances of C^{α} distances, variances of torsion angles or solvent-accessible surface areas. The training data for the study was collected from the BMRB and PDB databases, although the scarcity of order parameter depositions only allowed acquisition of data associated with 36 proteins that the authors split into training (26) and test sets (10). The authors compared their method to two baseline approaches, demonstrating superior prediction accuracy over a weighted contact-number model (WCN [112]), and comparable performance to a composite dynamic score with AlphaFold2 (cdsAF2 [113]).

Another machine learning approach, PDBCor [114], complements the above work by allowing detection of allostery and correlated motions in protein structure ensembles, implicitly involving dynamics at longer time scales. In this study, the authors represented protein models with either distance matrices that store pairwise distances between residue centroids, or angular matrices that store torsion angles. The underlying machine learning approach remained identical, regardless of the selected structure representation. The basic idea is that each row in the distance matrix forms a unique fingerprint of a residue by storing its distances to all other residues in the sequence. As multiple models are available in the NMR protein ensemble, multiple instances of the same residue vector can be compared and clustered with a Gaussian Mixture Model (GMM), grouping them into the expected number of protein states. Having such a clustering for each residue and each model in the protein ensemble, the algorithm calculates an adjusted mutual information for residue pairs. Mutual information is a statistical measure that quantifies the reduction in uncertainty about one variable given knowledge of another—in this case, how much knowing the conformational cluster of one residue tells us about another. Adjusted mutual information extends this concept by correcting for the amount of agreement expected by chance, resulting in a normalized score between 0 (no more agreement than random) and 1 (perfect correlation). Intuitively, if two residues have strongly correlated motion, knowing the state of the first allows one to infer the state of the second residue, with certainty or high probability. On the other hand, if there is no correlated motion, knowing the state of the first residue provides no information about the second. The authors used several protein systems to study the performance of the method, demonstrating its ability to detect both local and global system correlations. An interesting feature of the approach is that it does not rely on structure superposition, which can impact the assessment of correlated motion.

4. Spectrum reconstruction and quality enhancement

NMR experiments typically exploit the time evolution of nuclear spin coherences, and, in multiple pulse experiments, interconversions between different coherences. The latter produces Free Induction Decays (FIDs), which are conventionally transformed into the frequency domain using a Fourier transform—an information-preserving, bijective process—yielding multidimensional NMR spectra. Bottlenecks therein are the minimum measurement time that increases by 1–2 orders of magnitude per additional dimension (for example, a standard 3D NMR experiment takes about 1 day of measurement), the low signal-to-noise ratio of NMR resulting from the minute Boltzmann-weighted polarization of the nuclear spins, and the complex interaction network of the recorded spins that may yield peak splittings due to scalar couplings. The theory of NMR is well understood and allows for quantitatively accurate simulations of spectra, opening avenues for deep learning in spectrum reconstruction and quality enhancement (Table 4).

4.1. NUS reconstructions

Non-uniform sampling (NUS) enables significantly faster acquisition of multidimensional data compared to the conventional approach by leveraging sparse measurements in indirect time dimensions. The accuracy of the resulting NMR spectrum depends on the sampling schedule used to generate the sparse data and the algorithm used for data reconstruction. Commonly used methods for NUS reconstruction include iterative soft-thresholding [121,122] and sparse multidimensional iterative lineshape-enhanced reconstruction [123]. Supervised deep neural networks represent an alternative to existing algorithmic approaches.

Qu et al. [115] proposed an approach that employs an iterative framework in which undersampled NUS data first undergoes artifact suppression via a dense convolutional neural network. The resulting intermediate reconstruction is subsequently refined to ensure consistency with the original NUS signal. This iterative process—alternating between CNN-based artifact removal and consistency enforcement—is repeated to progressively improve spectral reconstruction. The model was trained using fully sampled synthetic time-domain NMR signals, which were subsequently subsampled according to a Poisson gap scheme to simulate realistic NUS conditions. The method was evaluated on experimentally acquired 2D and 3D spectra of small proteins, demonstrating performance comparable to state-of-the-art low-rank reconstruction methods while achieving about an order-of-magnitude reduction in computational time. Notably, the deep learning-based approach exhibited superior reconstruction fidelity at lower NUS densities (<20%). To optimize model performance, distinct network instances were trained separately for 2D and 3D spectra.

In another study, a deep Encoder-Decoder High-Resolution Network (EDHRN) has been introduced [116]. Drawing inspiration from the High-Resolution Net architecture [117] (a variant of CNN), originally

Table 4
Highlighted studies on ML approaches for spectrum reconstruction and quality enhancement.

Method	Model/approach	Dataset	Prediction scope
Qu [115] (2020)	Deep neural network Dense convolutional neural network	Simulated dataset with Poisson-gap sampling	NUS reconstruction
EDHRN [116] (2020)	Deep neural network Architecture developed based on high-resolution net (HRNet) [117] that consists of stacked encoder-decoder blocks arranged in two parallel subnetworks.	Simulated dataset withe xponential and Poisson-gap sampling	NUS reconstruction
FID-Net [3] (2020)	Deep neural network Convolutional neural network with dilated convolutions and gated activation units, architecture inspired by WaveNet [118]	Simulated dataset with Poisson-gap and random sampling	NUS reconstruction, Virtual decoupling of $^{13}C_{\alpha}$–$^{13}C_{\beta}$ in HNCA and HN(CO)CA spectra
Karunanithy [4] (2021)		Three simulated datasets, each for different decoupling task	Virtual decoupling of ^{13}CO–N spectra, ^{13}C–^{13}C correlation spectra of protein side chain, and $^{13}C_{\alpha}$–^{13}CO protein correlation spectra.
DN-UNet [2] (2021)	Deep neural network Convolutional neural network, variant of U-Net architecture. Size of the architecture was adjusted to the dimensionality of NMR spectra, yielding 12 (1D) and 9 (2D, 3D) convolutional and corresponding deconvolutional layers.	Simulated 1D spectra with M-to-S training examples, where multiple noisy spectra were formed out of each simulated spectrum.	Denoising of solution-state NMR spectra
PIPNet [119] (2023)	Deep neural network Ensemble of 16 6-layer convolutional LSTMs	Simulated dataset of 1D isotropic spectra and spectra acquired as different spinning rates.	Predicting isotropic proton spectra from sequence of MAS spectra acquired at different spinning rates.
AC-ResNet [5] (2023)	Deep neural network 72-layer variant of ResNet architecture that introduces addition and concatenation operations to reduce overfitting and peak misclassification.	Simulated dataset of real-time ZS spectra and ideal pure shift spectra.	Quality enhancement (denoising, linewidth reduction) of pure shift NMR spectra acquired with Zangger-Sterk pulse sequence.
SENNet [120] (2024)	Deep neural network Convolutional neural network, variant of U-Net architecture.	Simulated spectra consisting of small-molecule and macromolecule signals.	Separation of signals of small and macromolecules in ^{1}H NMR spectra of plasma and serum

developed for human pose estimation in computer vision, EDHRN features a distinctive structure comprising two parallel subnetworks. This design is particularly effective in reconstructing weak NMR signals. Each subnetwork consists of convolutional layers and encoder-decoder blocks, where data dimensionality is reduced through convolutional layers and subsequently restored via deconvolutional layers. The subnetworks operate at different resolutions, facilitating a dynamic exchange of information: activations from the upper subnetwork are down-sampled and fed into the lower subnetwork, while activations from the lower subnetwork are up-sampled and integrated into the upper subnetwork. This bidirectional flow enhances the network's ability to capture and reconstruct intricate spectral features. For training, the authors generated synthetic datasets by simulating fully sampled spectra, which were then undersampled using exponential and Poisson gap sampling schemes to mimic NUS conditions. The performance of EDHRN was evaluated on both two-dimensional and three-dimensional NMR spectra.

FID-Net [3] is a deep learning model designed for NUS reconstruction and virtual decoupling of $^{13}C^{\alpha}$–$^{13}C^{\beta}$ couplings in HNCA and HNcoCA experiments. The model takes as input sparse NMR data that has been Fourier-transformed in the direct dimension for HSQC experiments or in the ^{1}H,^{15}N-plane for HNCA and HNcoCA experiments. The architecture of FID-Net draws inspiration from WaveNet [118], a generative model originally developed for raw audio processing. It consists of stacked blocks of convolutional layers with varying dilation rates, enabling the model to capture both local and non-local dependencies within the input data. These dilated convolutional blocks are equipped with Gated Activation Units (GAUs) and residual connections. The GAUs play a crucial role in regulating the flow of information through the network, enhancing its ability to focus on relevant spectral features while mitigating noise. The output of convolution layers within blocks are aggregated and passed through a prediction head composed of two convolutional layers to generate the final reconstruction. FID-Net was trained exclusively on simulated data and evaluated on both simulated and experimental datasets. When compared to established NUS reconstruction methods such as SMILE [123] and hmsIST [122], FID-Net demonstrated comparable performance on high-quality simulated data. Notably, its reconstruction quality surpassed that of reference methods as the quality of the input data deteriorated, a condition simulated by introducing Gaussian noise with variable standard deviation.

4.2. *Virtual decoupling*

Strong dipolar interactions in solid state NMR, as well as unsuppressed scalar interactions in solution state NMR, can make NMR spectra difficult to interpret. Specifically, in a dense ^{1}H network, strong ^{1}H–^{1}H dipolar interactions in the solid state require both homo- and heteronuclear dipolar decoupling to convert broad signals into well-resolved peaks. Scalar interactions, particularly visible in liquid-state NMR spectra, carry valuable structural information but can also hinder accurate peak identification. For these reasons, decoupling and subsequent recoupling remain fundamental steps in NMR spectrum acquisition. A limitation of ^{13}C-detected protein NMR experiments is the presence of one-bond ^{13}C–^{13}C coupling evolution during acquisition, with coupling constants ranging from 35 to 55 Hz. If resolved, these significantly increase the number of peaks in the final spectrum, or, if unresolved, they contribute to peak broadening. Since this coupling is homonuclear, implementing experimental decoupling is less straightforward than for heteronuclear decoupling, and often introduces significant artifacts. As a result, virtual decoupling presents a more convenient alternative.

In an FID-Net follow-up study [4] three new instances of the FID-Net architecture were trained to address decoupling of 2D ^{13}C–^{13}C correlation spectra of protein side chains, CON (^{13}CO–^{15}N) spectra, and $^{13}C^{\alpha}$–^{13}CO protein correlation spectra. Like the original FID-Net model, these networks were trained exclusively on synthetic data and subsequently validated on experimentally acquired datasets. This approach demonstrated the versatility of FID-Net in adapting to various NMR applications. The retrained models offer an advantage over traditional experimental methods such as IPAP [124] and S^3E [125], which typically require the acquisition of multiple spectra to achieve effective decoupling. In contrast, FID-Net can perform these tasks using only a single spectrum, thereby reducing measurement time.

The resolution of proton solid-state NMR spectra is often limited by spectral broadening caused by strong dipolar interactions between nuclear spins. While magic-angle spinning (MAS) significantly reduces

these interactions, even the highest spinning rates achievable with current technology cannot fully eliminate interproton dipolar couplings. To address this limitation, the authors of PIPNet [119] proposed a deep learning-based approach designed to infer isotropic 1D NMR spectra, representing the spectra that would theoretically be obtained at infinite spinning rates. This method uses a variant of a convolutional long short-term memory (LSTM, Table 1) network, trained on millions of synthetic datasets, to reconstruct high-resolution spectra from experimental data collected at lower MAS frequencies. The model operates sequentially, where at each prediction step, the LSTM processes one spectrum—starting from the lowest MAS rate—and progressively refines the prediction of the isotropic spectrum. To estimate the uncertainty associated with the predictions, the authors built an ensemble of 16 identical LSTM architectures, each trained independently on distinct synthetic datasets. The final isotropic spectrum is obtained by averaging the outputs of all 16 models, while the standard deviation among them provides an estimation of the prediction uncertainty. Beyond its primary application in predicting isotropic spectra, the model also demonstrated utility in reconstructing a high-resolution spectrum (comparable to those obtained at high MAS rates), using a few spectra collected at lower spinning frequencies. This capability has practical implications, as it enables researchers to achieve spectral resolution traditionally limited to fast MAS conditions (>100 kHz) using data acquired at more accessible MAS rates (30–50 kHz). The model was applied to MAS spectra from eight different molecular solids, yielding resolution improvements with linewidths as narrow as 57 Hz—surpassing the resolution limits of the highest MAS rates currently available. Notably, even with modest MAS rates, the model produced spectra with linewidths comparable to fast MAS, thus opening new possibilities for obtaining well-resolved spectra on standard hardware, such as 1.3 mm MAS probes, without the need for specialized fast-spinning equipment. In a subsequent study, PIPNet2D [126], the approach was extended to 2D spectra.

4.3. Denoising

Noise is an inherent challenge in NMR spectroscopy, arising primarily from physical factors such as thermal fluctuations and imperfections of the instrument hardware. Inadequate signal averaging due to limited acquisition time can lead to a lower signal-to-noise ratio, making it harder to detect weak signals, reducing spectral resolution, and complicating peak identification—ultimately impacting the accuracy of data interpretation. Traditional noise reduction techniques often involve trade-offs between preserving weak signals and suppressing noise. More recently, machine learning-based approaches, particularly deep neural networks, have shown promise in denoising NMR spectra while preserving fine spectral details, making them a valuable tool for improving spectral quality in both solution and solid-state NMR applications.

Real-time experiments in NMR spectroscopy play a crucial role in monitoring dynamic processes and following chemical reactions. Pure shift methods [127] use homonuclear broadband decoupling to enhance spectral resolution by collapsing multiplets into singlets. However, techniques such as Zangger-Sterk (ZS) and PSYCHE suffer from sensitivity losses, often reaching one or two orders of magnitude. An Added and Concatenated Residual Network (AC-ResNet) [5] was designed to enhance the quality of pure shift NMR spectra acquired using the real-time ZS pulse sequence. The model employs a custom loss function, SM-CDMANE (Signal Matrix Matched Continuously Differentiable Mean Absolute Normalized Error), which increases its sensitivity to weak signals in the training process. The model was trained exclusively on simulated data and subsequently evaluated against experimentally acquired 1D NMR spectra, including both conventional and high-scan pure shift ZS spectra. AC-ResNet demonstrated strong performance in restoring high-quality spectra from noisy low-scan experiments, making it a tool for enhancing sensitivity, removing noise, reducing line widths and removing artifacts. In comparative analysis, the authors benchmarked their 72-layer AC-ResNet model against standard ResNet architectures with varying depths (36, 72, and 108 layers), providing justification for their architectural modifications and demonstrating the effectiveness of the approach.

DN-UNet [2] is a deep learning model designed for denoising 1D and multidimensional NMR spectra. The proposed model architecture is an adaptation of U-Net [128] (a type of CNN), which was originally proposed for medical image segmentation and has since become one of the most popular choices in applied machine learning in this area. The core concept behind the U-Net architecture is its ability to process entire images as input, systematically extracting high-level features through a series of convolution and pooling layers. These layers progressively reduce the spatial dimensions of the data, compressing the input into a compact representation at the network's bottleneck. This bottleneck serves as a compact summary of the entire input, retaining the most critical information necessary for reconstruction. Following the bottleneck, DN-UNet employs a sequence of deconvolutional (or up-sampling) layers designed to reconstruct the denoised output from bottleneck representation. The up-sampling pathway is interconnected with the initial down-sampling layers through residual connections, which help maintain the flow of information and preserve fine-grained details from the original input. The authors utilize the 1D architecture to handle data of higher dimensionalities by treating higher-dimensional data as a collection of 1D slices along the indirect dimensions. For 1D denoising tasks, the model incorporates 12 convolutional and deconvolutional layers, while for 2D and 3D data, the architecture is adjusted to use 9 layers. The model was trained with supervised learning using simulated spectra and M-to-S approach, which implies that several noisy examples were associated with each ground truth spectrum. Model performance was evaluated using pairs of spectra obtained under varying conditions, including different numbers of scans and sample concentrations. Notably, DN-UNet demonstrated denoising capabilities on spectra of organic compounds recorded with as few as four scans—compared to the 1024 scans used as a reference—or at concentrations 25 times lower than reference (0.02 mol/l versus 0.5 mol/l). When compared to its predecessor, NASR [129], DN-UNet exhibited improvements by denoising weak signals with intensities close to the noise amplitude while minimizing the introduction of false-positive signals. Overall, the authors demonstrate that DN-UNet can enhance the signal-to-noise ratio (SNR) of NMR spectra more than 200-fold in favourable situations.

In addition, new methods for NMR spectra quality enhancement come from the medical domain, such as metabolomics and MRS spectroscopy. For instance, in a study by Xiao et al., a method called Spectral Editing Neural Network (SENNet) [120] is presented to edit and separate the spectral signals of small and macromolecules in ^{1}H NMR spectra of serum and plasma, based on their peak linewidth. The model is a 1D variant of U-Net established with simulated spectra. Principal component analysis confirms that the spectra extracted by the model closely resemble those obtained through traditional NMR spectral editing methods, such as relaxation-edited and diffusion-edited spectra. In a similar way, proton magnetic resonance spectroscopy (^{1}H-MRS) enables the *in vivo* quantification of brain metabolites. MRS works by measuring the spectra of specific metabolites, such as *N*-acetylaspartate (NAA), glutamate, myo-inositol, and γ-aminobutyric acid (GABA), providing a metabolic fingerprint of different brain regions. *In vivo* brain spectra are typically degraded by low signal-to-noise ratio and line broadening. Therefore, convolutional neural networks (CNN) find applications in quantification of brain metabolites [130].

5. Conclusions and outlook

Machine learning is expected to have a profound impact on NMR spectroscopy, transforming both experimental workflows and data analysis, yet its full potential has not yet been realized. The rapid development of new deep learning models continues to expand the scope of ML-driven applications in NMR. Beyond ML architectures adapted by the NMR community, techniques such as vision transformers

[131] and diffusion models [132] hold promise for further advances. Vision transformers, which have demonstrated excellent performance in computer vision tasks by capturing long-range dependencies in image data, could be adapted for pattern recognition in multidimensional NMR spectra, peak classification, and spectral deconvolution. Similarly, diffusion models, which have gained traction in fields such as molecular design and image synthesis, could provide new ways to simulate realistic NMR spectra and improve spectral reconstruction from highly sparse datasets. However, challenges such as data scarcity, model generalizability, and the interpretability of deep learning predictions must still be addressed to fully unlock the transformative potential in NMR spectroscopy.

5.1. ML adoption in laboratory practice

The integration of deep learning models into research workflows extends beyond model development—it also requires effective adaptation, deployment, and accessibility to ensure widespread adoption within the scientific community. A key challenge in this process is the heterogeneity of NMR data, which can vary significantly between studied systems, NMR spectrometers, laboratories, and experimental setups, complicating the application of trained models. Additionally, the proliferation of fragmented, standalone ML solutions creates a maintenance burden, requiring researchers to allocate resources to software updates, format conversions and computational infrastructure. To address these challenges, platform-based solutions, such as NMRtist [44] and NMRBox [132], have emerged to centralize deep learning tools, standardize data handling, and reduce the technical overhead.

5.2. Reducing experimental input

One of the aims of AI in protein NMR spectroscopy is to reduce both human time involvement and NMR recording time for chemical shift assignment, structure determination and dynamics elucidation to 24 h, focusing for example on 2D homonuclear ^{1}H–^{1}H NOESY and ^{1}H–^{1}H TOCSY/COSY experiments, if necessary using ultra high fields (currently up to 1.2 GHz) or/and super resolution (SR) NMR [133]. Such an approach would allow also for straightforward sophisticated 2D NMR measurements on proteins expressed from eucaryotic sources.

5.3. NMR experiment design optimization for deep learning analysis

In contrast to the past when NMR pulse sequence development focused on the requirements of human-based analysis, it is now conceivable to design NMR pulse sequences that are optimized for analysis by deep learning algorithms. This approach not only overrides human limitations (for example to work on spectra with well-defined correlations such as frequencies instead of analyzing the FIDs) but enables evaluation of the power of a newly developed NMR experiments by AI-based analysis tools using simulated data, as has been done recently for the implementation of super resolution NMR [133] into the HNCA triple resonance experiment. In addition, such approaches may allow for a dynamic interplay between the NMR spectrometer and the analysis software (e.g., targeted acquisition [134]), continuing acquisition until the required spectral features have been resolved or the software determines that no further data is needed—thereby reducing the measurement time or yielding higher quality data.

5.4. Spectra databases

In contrast to the 14,500 NMR structures deposited in the PDB and a similar number of chemical shift depositions in the BMRB that provide the basis for chemical shift predictions and secondary structure determination, multi-dimensional NMR time-domain data or spectra of proteins are so far deposited very rarely (less than 1.5% of the BMRB entries) challenging the progress of deep learning applied to NMR significantly. Within this context it is remarkable that ARTINA was trained with experimental spectra data sets of only 100 proteins [6], which were made publicly available [67]. To facilitate further progress in AI-NMR it is crucial that more NMR datasets are deposited. For this purpose, the NMRPrime database was established [135], and NMR time-domain data can be uploaded also to the BMRB [136].

5.5. Toward comprehensive structure and dynamic determination

The NMR signature of a protein yields thousands of NMR probes of distinct character such as chemical shifts, NOEs, scalar couplings, transverse-relaxation rates through linewidth, etc. within a few 2D or 3D NMR experiments. With the help of machine learning it is the goal is to find a structural and dynamical representation of a protein consistent with the NMR probe data collected.

Glossary of abbreviations

The glossary lists abbreviations commonly used in machine learning and NMR spectroscopy. Abbreviations introduced exclusively in specific studies (e.g., model or software names) are not included.

BMRB	Biological magnetic resonance data Bank
CNN	Convolutional neural network
CSD	Cambridge structural database
DFT	Density functional theory
EM	Electron microscopy
eNOE	Exact nuclear Overhauser enhancement
FID	Free induction decay
GCNN	Graph convolutional neural network
GMM	Gaussian mixture model
GNN	Graph neural network
GPR	Gaussian process regression
LSTM	Long short-term memory
MAE	Mean absolute error
ML	Machine learning
NUS	Non-uniform sampling
PDB	Protein data Bank
RDC	Residual dipolar coupling
RMSD	Root-Mean-Square deviation
RMSE	Root-Mean-Square error
RNN	Recurrent neural network
SOAP	Smooth overlap of atomic positions
SR	Super resolution

CRediT authorship contribution statement

Piotr Klukowski: Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Roland Riek:** Writing – review & editing, Supervision, Conceptualization. **Peter Güntert:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

P.K. acknowledges Swiss National Science Foundation (SNSF) funding (10001393). R.R. acknowledges internal ETH funding. P.G. acknowledges funding by a Grant-in-Aid for Scientific Research (JP23K05660) from the Japan Society for the Promotion of Science (JSPS).

Data availability

No data was used for the research described in the article.

References

[1] K. Kazimierczuk, V. Orekhov, Non-uniform sampling: post-Fourier era of NMR data collection and processing, Magn. Reson. Chem. 53 (2015) 921–926, https://doi.org/10.1002/MRC.4284.

[2] K. Wu, J. Luo, Q. Zeng, X. Dong, J. Chen, C. Zhan, Z. Chen, Y. Lin, Improvement in signal-to-noise ratio of liquid-state NMR spectroscopy via a deep neural network DN-Unet, Anal. Chem. 93 (2021) 1377–1382, https://doi.org/10.1021/acs.analchem.0c03087.

[3] G. Karunanithy, D.F. Hansen, FID-net: a versatile deep neural network architecture for NMR spectral reconstruction and virtual decoupling, J. Biomol. NMR 75 (2021) 179–191, https://doi.org/10.1007/s10858-021-00366-w.

[4] G. Karunanithy, H.W. Mackenzie, D.F. Hansen, Virtual Homonuclear decoupling in direct detection nuclear magnetic resonance experiments using deep neural networks, J. Am. Chem. Soc. 143 (2021) 16935–16942, https://doi.org/10.1021/jacs.1c04010.

[5] Z. Yang, X. Zheng, X. Gao, Q. Zeng, C. Yang, J. Luo, C. Zhan, Y. Lin, Deep learning methodology for obtaining ultraclean pure shift proton nuclear magnetic resonance spectra, J. Phys. Chem. Lett. 14 (2023) 3397–3402, https://doi.org/10.1021/acs.jpclett.3c00455.

[6] P. Klukowski, R. Riek, P. Güntert, Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA, Nat. Commun. 13 (2022) 6151, https://doi.org/10.1038/s41467-022-33879-5.

[7] D.W. Li, A.L. Hansen, C. Yuan, L. Bruschweiler-Li, R. Brüschweiler, DEEP picker is a DEEP neural network for accurate deconvolution of complex two-dimensional NMR spectra, Nat. Commun. 12 (2021) 5229, https://doi.org/10.1038/s41467-021-25496-5.

[8] 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, D. Wilhelm, Deconvolution of 1D NMR spectra: a deep learning-based approach, J. Magn. Reson. 347 (2023) 107357, https://doi.org/10.1016/j.jmr.2022.107357.

[9] H. Wetton, P. Klukowski, R. Riek, P. Güntert, Chemical shift transfer: an effective strategy for protein NMR assignment with ARTINA, Front. Mol. Biosci. 10 (2023) 1244029, https://doi.org/10.3389/fmolb.2023.1244029.

[10] Y. Zhang, Z. Zhang, Y. Kagaya, G. Terashi, B. Zhao, Y. Xiong, D. Kihara, Distance-AF: Modifying Predicted Protein Structure Models by Alphafold2 with User-Specified Distance Constraints, BioRxiv (2023), https://doi.org/10.1101/2023.12.01.569498, 2023.12.01.569498.

[11] S. Kang, Y. Kwon, D. Lee, Y.-S. Choi, Predictive modeling of NMR chemical shifts without using atomic-level annotations, J. Chem. Inf. Model. 60 (2020) 3765–3769, https://doi.org/10.1021/acs.jcim.0c00494.

[12] Y. Guan, S.V. Shree Sowndarya, L.C. Gallegos, P.C. St, R.S. Paton John, Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network, Chem. Sci. 12 (2021) 12012–12026, https://doi.org/10.1039/d1sc03343c.

[13] E. Jonas, S. Kuhn, Rapid prediction of NMR spectral properties with quantified uncertainty, J. Chem. 11 (2019) 50, https://doi.org/10.1186/s13321-019-0374-3.

[14] F.M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, L. Emsley, Chemical shifts in molecular solids by machine learning, Nat. Commun. 9 (2018) 4501, https://doi.org/10.1038/s41467-018-06972-x.

[15] S. Liu, J. Li, K.C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima, T. Head-Gordon, Multiresolution 3D-DenseNet for chemical shift prediction in NMR crystallography, J. Phys. Chem. Lett. 10 (2019) 4558–4565, https://doi.org/10.1021/acs.jpclett.9b01570.

[16] M. Cordova, E.A. Engel, A. Stefaniuk, F. Paruzzo, A. Hofstetter, M. Ceriotti, L. Emsley, A machine learning model of chemical shifts for chemically and structurally diverse molecular solids, J. Phys. Chem. C 126 (2022) 16710–16720, https://doi.org/10.1021/acs.jpcc.2c03854.

[17] J. Li, K.C. Bennett, Y. Liu, M.V. Martin, T. Head-Gordon, Accurate prediction of chemical shifts for aqueous protein structure on "real world" data, Chem. Sci. 11 (2020) 3180–3191, https://doi.org/10.1039/c9sc06561j.

[18] A.L. Ptaszek, J. Li, R. Konrat, G. Platzer, T. Head-Gordon, UCBShift 2.0: bridging the gap from backbone to side chain protein chemical shift prediction for protein structures, J. Am. Chem. Soc. 146 (2024) 31733–31745, https://doi.org/10.1021/jacs.4c10474.

[19] Z. Yang, M. Chakraborty, A.D. White, Predicting chemical shifts with graph neural networks, Chem. Sci. 12 (2021) 10802–10809, https://doi.org/10.1039/d1sc01895g.

[20] J. Kirkwood, D. Hargreaves, S. O'Keefe, J. Wilson, Analysis of crystallization data in the protein data Bank, Acta Crystallogr. F 71 (2015) 1228–1234, https://doi.org/10.1107/s2053230x15014892.

[21] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R. M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G. N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr. D 67 (2011) 235–242, https://doi.org/10.1107/s0907444910045749.

[22] E. Callaway, Revolutionary cryo-EM is taking over structural biology, Nature 578 (2020) 201, https://doi.org/10.1038/d41586-020-00341-9.

[23] M. Peplow, Cryo-electron microscopy reaches resolution milestone, ACS Cent. Sci. 6 (2020) 1274–1277, https://doi.org/10.1021/acscentsci.0c01048.

[24] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[25] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A.J. Ballard, J. Bambrick, S.W. Bodenstein, D.A. Evans, C.C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A.I. Cowen-Rivers, A. Cowie, M. Figurnov, F.B. Fuchs, H. Gladman, R. Jain, Y.A. Khan, C.M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E.D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J.M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature 630 (2024) 493–500, https://doi.org/10.1038/s41586-024-07487-w.

[26] T.C. Terwilliger, D. Liebschner, T.I. Croll, C.J. Williams, A.J. McCoy, B.K. Poon, P. V. Afonine, R.D. Oeffner, J.S. Richardson, R.J. Read, P.D. Adams, AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination, Nat. Methods 21 (2023) 110–116, https://doi.org/10.1038/s41592-023-02087-4.

[27] A. Bax, G.M. Clore, Protein NMR: boundless opportunities, J. Magn. Reson. 306 (2019) 187–191, https://doi.org/10.1016/j.jmr.2019.07.037.

[28] M. Billeter, G. Wagner, K. Wüthrich, Solution NMR structure determination of proteins revisited, J. Biomol. NMR 42 (2008) 155–158, https://doi.org/10.1007/s10858-008-9277-8.

[29] H. Schwalbe, J. Winner, Protein Misfolding Disease: Overview of Liquid and Solid-State High Resolution NMR Studies, Modern Magnetic Resonance, Springer, in, 2008, pp. 1387–1391, https://doi.org/10.1007/1-4020-3910-7_152.

[30] K. Wüthrich, NMR with proteins and nucleic acids, Europhys. News 17 (1986) 11–13, https://doi.org/10.1051/epn/19861701011.

[31] R. Ishima, D.A. Torchia, Protein dynamics from NMR, Nat. Struct. Biol. 7 (2000) 740–743, https://doi.org/10.1038/78963.

[32] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, NMR evidence for slow collective motions in cyanometmyoglobin, Nat. Struct. Biol. 4 (1997) 292–297, https://doi.org/10.1038/nsb0497-292.

[33] L.E. Kay, NMR studies of protein structure and dynamics, J. Magn. Reson. 173 (2005) 193–207, https://doi.org/10.1016/j.jmr.2004.11.021.

[34] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, Biochemistry 28 (1989) 8972–8979, https://doi.org/10.1021/bi00449a003.

[35] R. Brüschweiler, R.R. Ernst, Molecular dynamics monitored by cross-correlated cross relaxation of spins quantized along orthogonal axes, J. Chem. Phys. 96 (1992) 1758–1766, https://doi.org/10.1063/1.462131.

[36] G.M. Clore, C.D. Schwieters, Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements, Biochemistry 43 (2004) 10678–10691, https://doi.org/10.1021/bi049357w.

[37] O.F. Lange, N.A. Lakomek, C. Farès, G.F. Schröder, K.F.A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B.L. De Groot, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, Science 320 (2008) 1471–1475, https://doi.org/10.1126/science.1157092.

[38] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science 278 (1997) 1111–1114, https://doi.org/10.1126/science.278.5340.1111.

[39] D. Strotz, J. Orts, H. Kadavath, M. Friedmann, D. Ghosh, S. Olsson, C.N. Chi, A. Pokharna, P. Güntert, B. Vögeli, R. Riek, Protein Allostery at atomic resolution, Angew. Chem. Int. Edit. 59 (2020) 22132–22139, https://doi.org/10.1002/anie.202008734.

[40] B. Vögeli, P. Güntert, R. Riek, Multiple-state ensemble structure determination from eNOE spectroscopy, Mol. Phys. 111 (2013) 437–454, https://doi.org/10.1080/00268976.2012.728257.

[41] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.

[42] S. Kuhn, H. Kolshorn, C. Steinbeck, N. Schlörer, Twenty years of nmrshiftdb2: a case study of an open database for analytical chemistry, Magn. Reson. Chem. 62 (2024) 74–83, https://doi.org/10.1002/mrc.5418.

[43] P. Klukowski, R. Riek, P. Güntert, Time-optimized protein NMR assignment with an integrative deep learning approach using AlphaFold and chemical shift prediction, Sci. Adv. 9 (2023) eadi9323, https://doi.org/10.1126/sciadv.adi9323.

[44] P. Klukowski, R. Riek, P. Guntert, NMRtist: an online platform for automated biomolecular NMR spectra analysis, Bioinformatics 39 (2023) btad066, https://doi.org/10.1093/bioinformatics/btad066.

[45] G.J. Kleywegt, R. Boelens, R. Kaptein, A versatile approach toward the partially automatic recognition of cross peaks in 2D 1H NMR spectra, J. Magn. Reson. 88 (1990) 601–608, https://doi.org/10.1016/0022-2364(90)90291-g.

[46] B.U. Meier, G. Bodenhausen, R.R. Ernst, Pattern recognition in two-dimensional NMR spectra, J. Magn. Reson. 60 (1984) 161–163, https://doi.org/10.1016/0022-2364(84)90043-x.

[47] P. Pfändler, B.U. Meier, R.R. Ernst, G. Bodenhausen, Toward automated assignment of nuclear magnetic resonance spectra: pattern recognition in two-dimensional correlation spectra, Anal. Chem. 57 (1985) 2510–2516, https://doi.org/10.1021/ac00290a018.

[48] R. Koradi, M. Billeter, M. Engeli, P. Güntert, K. Wüthrich, Automated peak picking and peak integration in macromolecular NMR spectra using AUTOPSY, J. Magn. Reson. 135 (1998) 288–297, https://doi.org/10.1006/jmre.1998.1570.
[49] D.M. Korzhnev, I.V. Ibraghimov, M. Billeter, V.Y. Orekhov, MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data, J. Biomol. NMR 21 (2001) 263–268, https://doi.org/10.1023/a:1012982830367.
[50] V.Y. Orekhov, I.V. Ibraghimov, M. Billeter, MUNIN: a new approach to multi-dimensional NMR spectra interpretation, J. Biomol. NMR 20 (2001) 49–60, https://doi.org/10.1023/a:1011234126930.
[51] B. Alipanahi, X. Gao, E. Karakoc, L. Donaldson, M. Li, PICKY: a novel SVD-based NMR spectra peak picking method, Bioinformatics 25 (2009) i268–i275, https://doi.org/10.1093/bioinformatics/btp225.
[52] S. Tikole, V. Jaravine, V. Rogov, V. Dötsch, P. Güntert, Peak picking NMR spectral data using non-negative matrix factorization, BMC Bioinformatics 15 (2014) 46, https://doi.org/10.1186/1471-2105-15-46.
[53] B.A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules, Methods Mol. Biol. 278 (2004) 313–352, https://doi.org/10.1385/1-59259-809-9:313.
[54] B.A. Johnson, R.A. Blevins, NMRView: A computer program for the visualization and analysis of NMR data, J. Biomol. NMR 4 (1994) 603–614, https://doi.org/10.1007/bf00404272.
[55] C. Bartels, T.H. Xia, M. Billeter, P. Güntert, K. Wüthrich, The program XEASY for computer-supported NMR spectral analysis of biological macromolecules, J. Biomol. NMR 6 (1995) 1–10, https://doi.org/10.1007/bf00417486.
[56] S.P. Skinner, R.H. Fogh, W. Boucher, T.J. Ragan, L.G. Mureddu, G.W. Vuister, CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis, J. Biomol. NMR 66 (2016) 111–124, https://doi.org/10.1007/s10858-016-0060-y.
[57] A. Rouh, A. Louis-Joseph, J.Y. Lallemand, Bayesian signal extraction from noisy FT NMR spectra, J. Biomol. NMR 4 (1994) 505–518, https://doi.org/10.1007/bf00156617.
[58] C. Antz, K.P. Neidig, H.R. Kalbitzer, A general Bayesian method for an automated signal class recognition in 2D NMR spectra combined with a multivariate discriminant analysis, J. Biomol. NMR 5 (1995) 287–296, https://doi.org/10.1007/bf00211755.
[59] Y. Cheng, X. Gao, F. Liang, Bayesian peak picking for NMR spectra, Genom. Proteom. Bioinf. 12 (2014) 39–47, https://doi.org/10.1016/j.gpb.2013.07.003.
[60] D.S. Garrett, R. Powers, A.M. Gronenborn, G.M. Clore, A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams, J. Magn. Reson. 95 (1991) 214–220, https://doi.org/10.1016/0022-2364(91)90341-p.
[61] Z. Liu, A. Abbas, B.-Y. Jing, X. Gao, WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering, Bioinformatics 28 (2012) 914–920, https://doi.org/10.1093/bioinformatics/bts078.
[62] J.M. Würz, P. Güntert, Peak picking multidimensional NMR spectra with the contour geometry based algorithm CYPICK, J. Biomol. NMR 67 (2017) 63–76, https://doi.org/10.1007/s10858-016-0084-3.
[63] S.A. Corne, A.P. Johnson, J. Fisher, An artificial neural network for classifying cross peaks in two-dimensional NMR spectra, J. Magn. Reson. 100 (1992) 256–266, https://doi.org/10.1016/0022-2364(92)90260-e.
[64] E.A. Carrara, F. Pagliari, C. Nicolini, Neural networks for the peak-picking of nuclear magnetic resonance spectra, Neural Netw. 6 (1993) 1023–1032, https://doi.org/10.1016/s0893-6080(09)80012-9.
[65] P. Klukowski, M. Augoff, M. Zięba, M. Drwal, A. Gonczarek, M.J. Walczak, NMRNet: a deep learning approach to automated peak picking of protein NMR spectra, Bioinformatics 34 (2018) 2590–2597, https://doi.org/10.1093/bioinformatics/bty134.
[66] P. Klukowski, M.J. Walczak, A. Gonczarek, J. Boudet, G. Wider, Computer vision-based automated peak picking applied to protein NMR spectra, Bioinformatics 31 (2015) 2981–2988, https://doi.org/10.1093/bioinformatics/btv318.
[67] P. Klukowski, F.F. Damberger, F.H.T. Allain, H. Iwai, H. Kadavath, T.A. Ramelot, G.T. Montelione, R. Riek, P. Güntert, The 100-protein NMR spectra dataset: a resource for biomolecular NMR data analysis, Sci. Data 11 (2024) 30, https://doi.org/10.1038/s41597-023-02879-5.
[68] E. Schmidt, P. Güntert, A new algorithm for reliable and general NMR resonance assignment, J. Am. Chem. Soc. 134 (2012) 12817–12829, https://doi.org/10.1021/ja305091n.
[69] A. Bahrami, A.H. Assadi, J.L. Markley, H.R. Eghbalnia, Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy, PLoS Comput. Biol. 5 (2009) e1000307, https://doi.org/10.1371/journal.pcbi.1000307.
[70] D.E. Zimmerman, C.A. Kulikowski, Y. Huang, W. Feng, M. Tashiro, S. Shimotakahara, C. Ya Chien, R. Powers, G.T. Montelione, Automated analysis of protein NMR assignments using methods from artificial intelligence, J. Mol. Biol. 269 (1997) 592–610, https://doi.org/10.1006/jmbi.1997.1052.
[71] Y.S. Jung, M. Zweckstetter, Mars - robust automatic backbone assignment of proteins, J. Biomol. NMR 30 (2004) 11–23, https://doi.org/10.1023/b:jnmr.0000042954.99056.ad.
[72] M.F. Hsu, M.K. Sriramoju, C.H. Lai, Y.R. Chen, J.S. Huang, T.P. Ko, K.F. Huang, S. T.D. Hsu, Structure, dynamics, and stability of the smallest and most complex 71 protein knot, J. Biol. Chem. 300 (2024) 105553, https://doi.org/10.1016/j.jbc.2023.105553.
[73] R.J. Flood, L. Cerofolini, M. Fragai, P.B. Crowley, Multivalent Calixarene complexation of a designed Pentameric lectin, Biomacromolecules 25 (2024) 1303–1309, https://doi.org/10.1021/acs.biomac.3c01280.
[74] A. Kumar, O. Vadas, N.D.S. Pacheco, X. Zhang, K. Chao, N. Darvill, H. Rasmussen, Y. Xu, G.M.H. Lin, F.A. Stylianou, J.S. Pedersen, S.L. Rouse, M.L. Morgan, D. Soldati-Favre, S. Matthews, Structural and regulatory insights into the glideosome-associated connector from toxoplasma gondii, Elife 12 (2023) e86049, https://doi.org/10.7554/elife.86049.
[75] S. McKenna, F. Aylward, X. Miliara, R.J. Lau, C.B. Huemer, S.P. Giblin, K.K. Huse, M. Liang, L. Reeves, M. Pearson, Y. Xu, S.L. Rouse, J.E. Pease, S. Sriskandan, T. F. Kagawa, J. Cooney, S. Matthews, The protease associated (PA) domain in ScpA from streptococcus pyogenes plays a role in substrate recruitment, Biochim. Biophys. Acta, Proteins Proteomics 1871 (2023) 140946, https://doi.org/10.1016/j.bbapap.2023.140946.
[76] L. Brigandat, M. Laux, C. Marteau, L. Cole, A. Böckmann, L. Lecoq, M.L. Fogeron, M. Callon, NMR side-chain assignments of the Crimean–Congo hemorrhagic fever virus glycoprotein n cytosolic domain, Magn. Reson. 5 (2024) 95–101, https://doi.org/10.5194/mr-5-95-2024.
[77] M. Gadanecz, Z. Fazekas, G. Pálfy, D. Karancsiné Menyhárd, A. Perczel, NMR-chemical-shift-driven protocol reveals the cofactor-bound, complete structure of dynamic intermediates of the catalytic cycle of oncogenic KRAS G12C protein and the significance of the Mg2+ ion, Int. J. Mol. Sci. 24 (2023) 12101, https://doi.org/10.3390/ijms241512101/s1.
[78] S. Yanaka, A. Kodama, S. Nishiguchi, R. Fujita, J. Shen, P. Boonsri, D. Sung, Y. Isono, H. Yagi, Y. Miyanoiri, T. Uchihashi, K. Kato, Identification of potential C1-binding sites in the immunoglobulin CL domains, Int. Immunol. 36 (2024) 405–412, https://doi.org/10.1093/intimm/dxae017.
[79] C.V. Holler, N.M. Petersson, M. Brohus, M.A. Niemelä, E.D. Iversen, M. T. Overgaard, H. Iwaï, R. Wimmer, Allosteric changes in protein stability and dynamics as pathogenic mechanism for calmodulin variants not affecting Ca2+ coordinating residues, Cell Calcium 117 (2024) 102831, https://doi.org/10.1016/j.ceca.2023.102831.
[80] J. Giudice, D.D. Brauer, M. Zoltek, A.L. Vázquez Maldonado, M. Kelly, A. Schepartz, Requirements for efficient endosomal escape by designed mini-proteins, BioRxiv (2024), https://doi.org/10.1101/2024.04.05.588336, 2024.04.05.588336.
[81] Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, J. Biomol. NMR 48 (2010) 13–22, https://doi.org/10.1007/s10858-010-9433-9.
[82] P. Klukowski, M. Augoff, M. Zamorski, A. Gonczarek, M.J. Walczak, Application of Dirichlet process mixture model to the identification of spin systems in protein NMR spectra, J. Biomol. NMR 71 (2018) 11–18, https://doi.org/10.1007/s10858-018-0185-2.
[83] S. Neal, A.M. Nip, H. Zhang, D.S. Wishart, Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts, J. Biomol. NMR 26 (2003) 215–240, https://doi.org/10.1023/a:1023812930288.
[84] Y. Shen, A. Bax, Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology, J. Biomol. NMR 38 (2007) 289–302, https://doi.org/10.1007/S10858-007-9166-6.
[85] M. Iwadate, T. Asakura, M.P. Williamson, C(α) and C(β) carbon-13 chemical shifts in proteins from an empirical database, J. Biomol. NMR 13 (1999) 199–211, https://doi.org/10.1023/a:1008376710086.
[86] K.J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, Fast and accurate predictions of protein NMR chemical shifts from interatomic distances, J. Am. Chem. Soc. 131 (2009) 13894–13895, https://doi.org/10.1021/ja903772t.
[87] D.W. Li, R. Brüschweiler, PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles, J. Biomol. NMR 54 (2012) 257–265, https://doi.org/10.1007/s10858-012-9668-8.
[88] J. Meiler, PROSHIFT: protein chemical shift prediction using artificial neural networks, J. Biomol. NMR 26 (2003) 25–37, https://doi.org/10.1023/a:1023060720156.
[89] B. Han, Y. Liu, S.W. Ginzinger, D.S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction, J. Biomol. NMR 50 (2011) 43–57, https://doi.org/10.1007/s10858-011-9478-4.
[90] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J. Biomol. NMR 44 (2009) 213–223, https://doi.org/10.1007/s10858-009-9333-z.
[91] Y. Shen, A. Bax, Prediction of Xaa-pro peptide bond conformation from sequence and chemical shifts, J. Biomol. NMR 46 (2010) 199–204, https://doi.org/10.1007/s10858-009-9395-y.
[92] H. Zhang, S. Neal, D.S. Wishart, RefDB: a database of uniformly referenced protein chemical shifts, J. Biomol. NMR 25 (2003) 173–195, https://doi.org/10.1023/a:1022836027055.
[93] D. Li, R. Brüschweiler, PPM-one: a static protein structure based chemical shift predictor, J. Biomol. NMR 62 (2015) 403–409, https://doi.org/10.1007/s10858-015-9958-z.
[94] A.P. Bartók, R. Kondor, G. Csányi, On representing chemical environments, Phys. Rev. B 87 (2013) 184115, https://doi.org/10.1103/physrevb.87.184115.
[95] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The Cambridge structural database, Acta Cryst. B 72 (2016) 171–179, https://doi.org/10.1107/s2052520616003954.
[96] W. Gerrard, L.A. Bratholm, M.J. Packer, A.J. Mulholland, D.R. Glowacki, C. P. Butts, IMPRESSION – prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy, Chem. Sci. 11 (2020) 508–515, https://doi.org/10.1039/c9sc03854j.
[97] M. Rupp, R. Ramakrishnan, O.A. Von Lilienfeld, Machine learning for quantum mechanical properties of atoms in molecules, J. Phys. Chem. Lett. 6 (2015) 3309–3313, https://doi.org/10.1021/acs.jpclett.5b01456.

[98] B. Huang, O.A. von Lilienfeld, Quantum machine learning using atom-in-molecule-based fragments selected on-the-fly, Nat. Chem. 12 (2017) 945–951, https://doi.org/10.1038/s41557-020-0527-z.
[99] F.A. Faber, A.S. Christensen, B. Huang, O.A. Von Lilienfeld, Alchemical and structural distribution based representation for universal quantum machine learning, J. Chem. Phys. 148 (2018) 241717, https://doi.org/10.1063/1.5020710/961132.
[100] C. Venkata, M.J. Forster, P.W.A. Howe, C. Steinbeck, The potential utility of predicted one bond carbon-proton coupling constants in the structure elucidation of small organic molecules by NMR spectroscopy, PloS One 9 (2014) e111576, https://doi.org/10.1371/journal.pone.0111576.
[101] S.G. Smith, J.M. Goodman, Assigning stereochemistry to single diastereoisomers by GIAO NMR calculation: the DP4 probability, J. Am. Chem. Soc. 132 (2010) 12946–12959, https://doi.org/10.1021/ja105035r.
[102] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: the human metabolome database for 2018, Nucleic Acids Res. 46 (2017) D608–D617, https://doi.org/10.1093/nar/gkx1089.
[103] K.T. Schütt, H.E. Sauceda, P.J. Kindermans, A. Tkatchenko, K.R. Müller, SchNet - a deep learning architecture for molecules and materials, J. Chem. Phys. 148 (2018) 241722, https://doi.org/10.1063/1.5019779/962591.
[104] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, International Conference on Learning Representations (2017), https://doi.org/10.48550/arXiv.1609.02907.
[105] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, https://doi.org/10.1109/cvpr.2017.243.
[106] CHESHIRE, Chemical Shift Repository. http://cheshirenmr.info, 2025 (accessed April 7, 2025).
[107] Predicting Molecular Properties | Kaggle, (n.d.), https://www.kaggle.com/competitions/champs-scalar-coupling (accessed January 17, 2025).
[108] R. Ramakrishnan, P.O. Dral, M. Rupp, O.A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Sci. Data 1 (2014) 140022, https://doi.org/10.1038/sdata.2014.22.
[109] L.A. Bratholm, W. Gerrard, B. Anderson, S. Bai, S. Choi, L. Dang, P. Hanchar, A. Howard, S. Kim, Z. Kolter, R. Kondor, M. Kornbluth, Y. Lee, Y. Lee, J.P. Mailoa, T.T. Nguyen, M. Popovic, G. Rakocevic, W. Reade, W. Song, L. Stojanovic, E. H. Thiede, N. Tijanic, A. Torrubia, D. Willmott, C.P. Butts, D.R. Glowacki, A community-powered search of machine learning strategy space to find NMR property prediction models, PloS One 16 (2021) e0253612, https://doi.org/10.1371/journal.pone.0253612.
[110] Q. Wang, Z. Miao, X. Xiao, X. Zhang, D. Yang, B. Jiang, M. Liu, Prediction of order parameters based on protein NMR structure ensemble and machine learning, J. Biomol. NMR 78 (2024) 87–94, https://doi.org/10.1007/s10858-024-00435-w.
[111] G. Lipari, A. Szabo, Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, J. Am. Chem. Soc. 104 (1982) 4546–4559, https://doi.org/10.1021/ja00381a009.
[112] D. Ming, R. Brüschweilery, Reorientational contact-weighted elastic network model for the prediction of protein dynamics: comparison with NMR relaxation, Biophys. J. 90 (2006) 3382–3388, https://doi.org/10.1529/biophysj.105.071902.
[113] P. Ma, D.W. Li, R. Brüschweiler, Predicting protein flexibility with AlphaFold, Proteins 91 (2023) 847–855, https://doi.org/10.1002/prot.26471.
[114] D. Ashkinadze, P. Klukowski, H. Kadavath, P. Güntert, R. Riek, PDBcor: An automated correlation extraction calculator for multi-state protein structures, Structure 30 (2022) 646–652.e2, https://doi.org/10.1016/j.str.2021.12.002.
[115] X. Qu, Y. Huang, H. Lu, T. Qiu, D. Guo, T. Agback, V. Orekhov, Z. Chen, Accelerated nuclear magnetic resonance spectroscopy with deep learning, Angew. Chem. Int. Edit. 59 (2020) 10297–10300, https://doi.org/10.1002/anie.201908162.
[116] J. Luo, Q. Zeng, K. Wu, Y. Lin, Fast reconstruction of non-uniform sampling multidimensional NMR spectroscopy via a deep neural network, J. Magn. Reson. 317 (2020) 106772, https://doi.org/10.1016/j.jmr.2020.106772.
[117] K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation, IEEE Conference on Computer Vision and Pattern Recognition (2019), https://doi.org/10.1109/cvpr.2019.00584.
[118] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K, Kavukcuoglu, WaveNet, A Generative Model for Raw Audio, 2016. https://arxiv.org/abs/1609.03499v2. (Accessed 18 April 2025).
[119] M. Cordova, P. Moutzouri, B. Simões de Almeida, D. Torodii, L. Emsley, Pure isotropic proton NMR spectra in solids using deep learning, Angew. Chem. Int. Edit. 62 (2023) e202216607, https://doi.org/10.1002/anie.202216607.
[120] X. Xiao, Q. Wang, X. Chai, X. Zhang, B. Jiang, M. Liu, Using neural networks to obtain NMR spectra of both small and macromolecules from blood samples in a single experiment, Commun. Chem. 7 (2024) 167, https://doi.org/10.1038/s42004-024-01251-x.
[121] A.S. Stern, D.L. Donoho, J.C. Hoch, NMR data processing using iterative thresholding and minimum l1-norm reconstruction, J. Magn. Reson. 188 (2007) 295–300, https://doi.org/10.1016/j.jmr.2007.07.008.
[122] S.G. Hyberts, A.G. Milbradt, A.B. Wagner, H. Arthanari, G. Wagner, Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson gap scheduling, J. Biomol. NMR 52 (2012) 315–327, https://doi.org/10.1007/s10858-012-9611-z.
[123] J. Ying, F. Delaglio, D.A. Torchia, A. Bax, Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data, J. Biomol. NMR 68 (2017) 101–118, https://doi.org/10.1007/s10858-016-0072-7.
[124] J.J. Chou, S. Gaemers, B. Howder, J.M. Louis, A. Bax, A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles, J. Biomol. NMR 21 (2001) 377–382, https://doi.org/10.1023/a:1013336502594.
[125] A. Meissner, J. Duus, O.W. Sørensen, Spin-State-Selective Excitation, Application for E.COSY-type measurement of JHH coupling constants, J. Magn. Reson. 128 (1997) 92–97, https://doi.org/10.1006/jmre.1997.1213.
[126] P. Moutzouri, M. Cordova, B. Simões de Almeida, D. Torodii, L. Emsley, Two-dimensional pure isotropic proton solid state NMR, Angew. Chem. Int. Edit. 62 (2023) e202301963, https://doi.org/10.1002/anie.202301963.
[127] K. Zangger, Pure shift NMR, Prog. Nucl. Magn. Reson. Spectrosc. 86–87 (2015) 1–20, https://doi.org/10.1016/j.pnmrs.2015.02.002.
[128] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Intervention (2015), https://doi.org/10.1007/978-3-319-24574-4_28.
[129] B. Jiang, F. Luo, Y. Ding, P. Sun, X. Zhang, L. Jiang, C. Li, X.A. Mao, D. Yang, C. Tang, M. Liu, NASR: an effective approach for simultaneous noise and artifact suppression in NMR spectroscopy, Anal. Chem. 85 (2013) 2523–2528, https://doi.org/10.1021/ac303726p.
[130] H.H. Lee, H. Kim, Intact metabolite spectrum mining by deep learning in proton magnetic resonance spectroscopy of the brain, Magn. Reson. Med. 82 (2019) 33–48, https://doi.org/10.1002/mrm.27727.
[131] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations, 2021. https://arxiv.org/abs/2010.11929v2 (accessed July 25, 2023).
[132] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: International Conference on Neural Information Processing Systems, 2020. https://arxiv.org/abs/2006.11239 (accessed August 8, 2024).
[133] L. Wenchel, O. Gampp, R. Riek, Super-resolution NMR spectroscopy, J. Magn. Reson. 366 (2024) 107746, https://doi.org/10.1016/J.JMR.2024.107746.
[134] V.A. Jaravine, V.Y. Orekhov, Targeted acquisition for real-time NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 13421–13426, https://doi.org/10.1021/ja062146p.
[135] Initiative for primary bio-NMR open research data | NMRprime, (n.d.), https://nmrprime.ethz.ch/ (accessed April 20, 2025).
[136] J.C. Hoch, K. Baskaran, H. Burr, J. Chin, H.R. Eghbalnia, T. Fujiwara, M.R. Gryk, T. Iwata, C. Kojima, G. Kurisu, D. Maziuk, Y. Miyanoiri, J.R. Wedell, C. Wilburn, H. Yao, M. Yokochi, Biological Magnetic Resonance Data Bank, Nucleic Acids Res. 51 (2023) D368–D376, https://doi.org/10.1093/nar/gkac1050.