Structure

Review

Biomolecular NMR spectroscopy in the era of artificial intelligence

Vaibhav Kumar Shukla,[1] Gabriella T. Heller,[1,*] and D. Flemming Hansen[1,*]
[1]Department of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, UK
*Correspondence: g.heller@ucl.ac.uk (G.T.H.), d.hansen@ucl.ac.uk (D.F.H.)
https://doi.org/10.1016/j.str.2023.09.011

SUMMARY

Biomolecular nuclear magnetic resonance (NMR) spectroscopy and artificial intelligence (AI) have a burgeoning synergy. Deep learning-based structural predictors have forever changed structural biology, yet these tools currently face limitations in accurately characterizing protein dynamics, allostery, and conformational heterogeneity. We begin by highlighting the unique abilities of biomolecular NMR spectroscopy to complement AI-based structural predictions toward addressing these knowledge gaps. We then highlight the direct integration of deep learning approaches into biomolecular NMR methods. AI-based tools can dramatically improve the acquisition and analysis of NMR spectra, enhancing the accuracy and reliability of NMR measurements, thus streamlining experimental processes. Additionally, deep learning enables the development of novel types of NMR experiments that were previously unattainable, expanding the scope and potential of biomolecular NMR spectroscopy. Ultimately, a combination of AI and NMR promises to further revolutionize structural biology on several levels, advance our understanding of complex biomolecular systems, and accelerate drug discovery efforts.

INTRODUCTION

Deep learning, a sub-field of machine learning and artificial intelligence (AI), has led to impressive advancements in many disciplines, particularly computer vision and language processing.[1–4] Due to the flexibility of deep learning, it is now deeply embedded into many aspects of science and daily life, including highly sophisticated large language models (LLMs), creation of digital art and music, as well as accurate predictions of protein structure from amino-acid sequences.[1–4]

Like numerous scientific disciplines, deep learning, both indirectly and directly, has revolutionized the field of biomolecular nuclear magnetic resonance (NMR) spectroscopy. Since its inception in the 1940s, NMR has undergone a dramatic expansion from its original application in materials science and chemistry to a versatile tool for environmental analysis, drug discovery, characterization of macromolecular complexes, and metabolic function in humans.[5] Biomolecular NMR spectroscopy, in particular, is a cornerstone in structural biology, providing atomic-level insights into the interactions and dynamics of biomolecules. Deep-learning based structure predictors like AlphaFold2[6] and ESM-2[7] have demonstrated remarkable capabilities for protein structure determination (Figure 1A), offering freely available predictions for entire proteomes, including the human proteome.[8] These tools and others have become invaluable aids to experimental structure determination methods, such as X-ray crystallography, cryo-electron microscopy (cryo-EM), and NMR spectroscopy, and are also playing increasingly important roles in drug discovery.[9–11] However, it is important to acknowledge that while these AI-based tools are highly useful, they do have certain limitations in providing insights into various aspects of protein biology including the mechanism of folding, stability, post-translational modifications, mutations, and intrinsically disordered proteins (IDPs). Additionally, these software tools are constrained in their capacity to predict conformational switching, as well as to provide detailed understanding of protein functional dynamics.[6,7]

Limitations in these AI-based tools have underscored the significance of NMR as an experimental tool in directly addressing these crucial knowledge gaps. In the first two sections of this review, we highlight how biomolecular NMR, particularly solution-state NMR methods characterizing protein dynamics, can effectively fill these gaps in our understanding, complementing the substantial progress achieved through structural predictors (Figure 1). In the final section of the review, we highlight how deep learning can be directly leveraged in the context of solution-state NMR to analyze spectra, improve the accuracy and reliability of NMR measurements, streamline the time and expenses associated with NMR experiments, and enable the development of novel types of NMR experiments that were previously unattainable (Figure 2).

NMR spectroscopy is founded on the principle that nuclei interact with a magnetic field, providing valuable information about their local chemical environments. When subjected to an external magnetic field, the nuclear spin magnetization resonates at a specific frequency (the Larmor frequency), which is influenced by factors such as the nucleus' gyromagnetic ratio, the magnetic field strength, and the chemical environment near the nucleus.
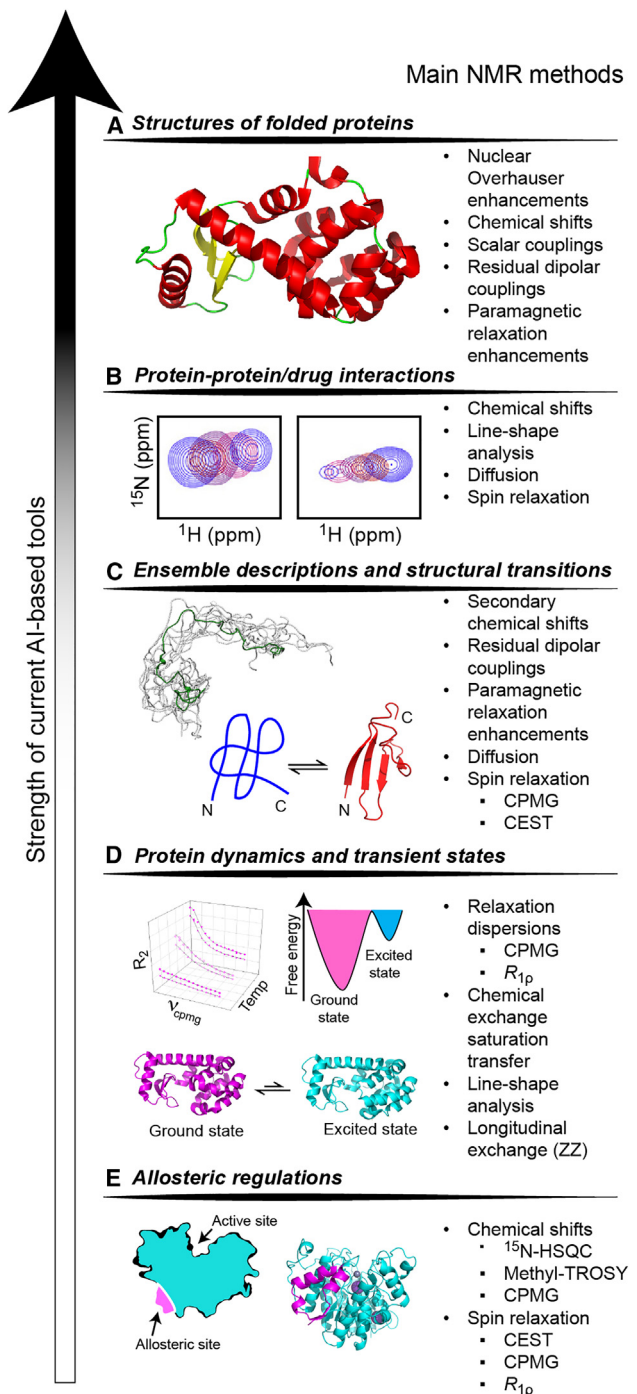
Main NMR methods

A *Structures of folded proteins*

- Nuclear Overhauser enhancements
- Chemical shifts
- Scalar couplings
- Residual dipolar couplings
- Paramagnetic relaxation enhancements

B *Protein-protein/drug interactions*

- Chemical shifts
- Line-shape analysis
- Diffusion
- Spin relaxation

C *Ensemble descriptions and structural transitions*

- Secondary chemical shifts
- Residual dipolar couplings
- Paramagnetic relaxation enhancements
- Diffusion
- Spin relaxation
  - CPMG
  - CEST

D *Protein dynamics and transient states*

- Relaxation dispersions
  - CPMG
  - $R_{1\rho}$
- Chemical exchange saturation transfer
- Line-shape analysis
- Longitudinal exchange (ZZ)

E *Allosteric regulations*

- Chemical shifts
  - ^{15}N-HSQC
  - Methyl-TROSY
  - CPMG
- Spin relaxation
  - CEST
  - CPMG
  - $R_{1\rho}$

Figure 1. Illustrating the spectrum of AI-based strengths in characterizing protein structure and dynamics along with the high complementarity of NMR spectroscopy

(A–E) AI methods have achieved remarkable success in predicting the structures of small proteins (A) and occasionally in predicting protein-ligand interactions (B). However, the characterization of heterogeneous conformational ensembles, such as those exhibited by IDPs and proteins undergoing significant structural transitions, poses an increasing challenge for AI methods (C). Furthermore, accurately describing protein dynamics and transient states (D) and capturing allosteric regulation (E) remain extremely difficult for current AI-based approaches. To address these challenges, we highlight the application of NMR techniques, which provide valuable experimental data on

The chemical shift, which relates to the Larmor frequency, is the easiest biomolecular NMR parameter to measure and it offers crucial biochemical insights. For instance, perturbations in the chemical shifts of backbone nuclei can indicate binding interactions or changes in secondary structure (Figure 1B). Through-space interactions can be investigated with the use of Nuclear Overhauser Effects (NOEs) for distances less than approximately 6 Å or paramagnetic relaxation enhancements (PREs) for distances between 10 and 25 Å.[12,13] Prior to the rise of highly accurate AI-based structural predictors for proteins, NMR-based structural determination posed several challenges. The complexity and interpretive difficulty of the resulting spectra, coupled with the indirect nature of the obtained parameters as reporters of structural information, often made the process both challenging and laborious, although attempts have been made to automate the process.[14] Moreover, NMR-based structural determination methods are primarily suitable for relatively small proteins (<40 kDa). Although NMR-based structural determination has to some extent been overshadowed by AI protein structure predictors, X-ray crystallography, and cryo-EM, NMR still provides invaluable insights where the predictive models fall short, particularly in areas concerning dynamics, rare folds, ligand binding, and alternative conformations.

Biomolecular NMR is often solely viewed as a structural technique, however, it is also uniquely able to characterize protein dynamics and flexibility across many timescales with atomic resolution (Figures 1C and 1D).[15,16] For example, some NMR methods, such as the classical model-free analyses[17] enable the study of motions on the picosecond (ps) to nanosecond (ns) timescale. For microsecond (μs) to millisecond (ms) motions, relaxation dispersion measurements, including $R_{1\rho}$ experiments,[18,19] Carr-Purcell-Meiboom-Gill (CPMG), chemical exchange saturation transfer (CEST), or dark state exchange saturation transfer (DEST) can be employed. Additionally, slower motions can be characterized using longitudinal exchange methods.[20]

The ability of NMR spectroscopy to characterize protein dynamics facilitates a comprehensive understanding of the behavior of proteins beyond static structures. By utilizing multiple nuclei as probes and operating under equilibrium conditions, without the need for external temperature or pressure changes, NMR can measure protein dynamics at the atomic level across a broad range of timescales. This dynamic and quantitative approach allows for the exploration of exchange dynamics between interconverting states, which is crucial for unraveling processes such as macromolecular recognition, allostery, signal transduction, and assembly. Notably, NMR observables provide valuable insights as they represent weighted time- and ensemble averages of proteins, highlighting the inherently dynamic nature of biomolecules, including protein motion, conformational sampling, and transient non-covalent interactions. In certain cases, NMR can also provide state-specific parameters that are challenging to acquire through alternative methods.

protein conformation, flexibility, and interactions. By combining the strengths of AI and NMR, a more comprehensive understanding of protein structures and dynamics can be achieved, offering the potential for new discoveries in the field of structural biology.
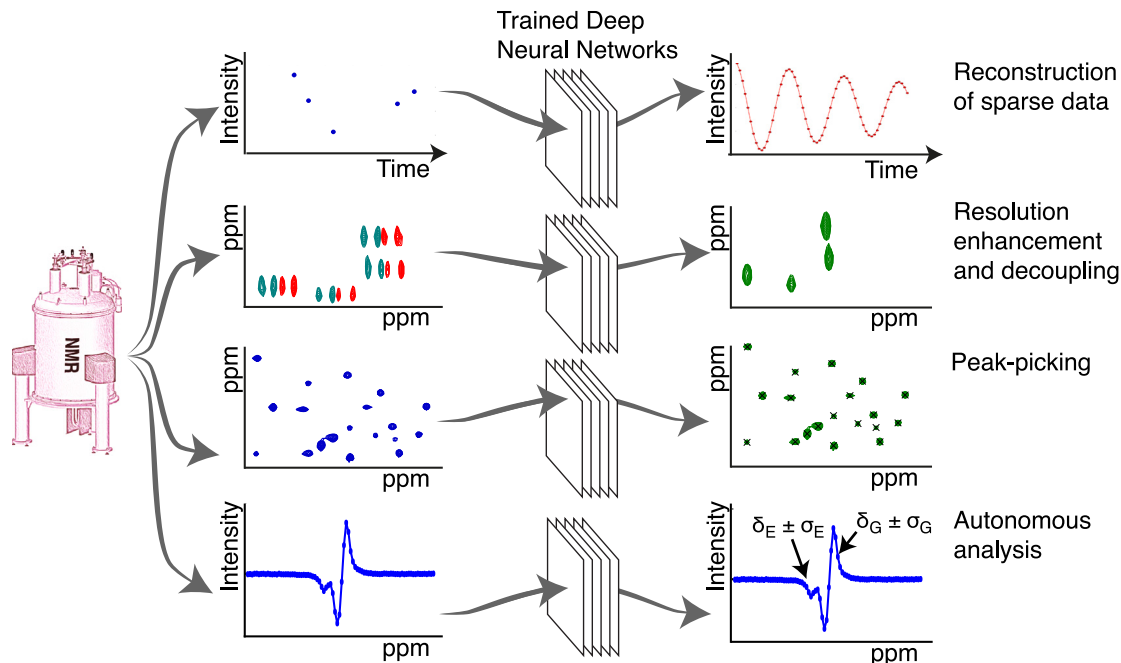
Structure
Review

CellPress
OPEN ACCESS

Figure 2. Direct application of AI-based methods to unleash the power of NMR spectroscopy
Illustration of the transformative power of DNNs to reconstruct sparce data, enhance resolution and decouple spectra with virtual homonuclear decouplings, pick peaks in crowded spectra, and perform autonomous analyses.

BIOMOLECULAR NMR AS A COMPLIMENTARY TOOL TO COMPUTATIONAL STRUCTURAL BIOLOGY, X-RAY, AND CRYO-EM

Characterization of protein dynamics by solution-state NMR

Protein structural dynamics is a crucial property of enzyme function, as highlighted by the induced fit model of enzyme kinetics,[21,22] which posits that enzymes undergo conformational changes upon binding to a substrate to achieve optimal catalytic activity. An alternative to this model is the conformational selection[23] or population shift model, which suggests that proteins can exist in thermally excited functional states, even without a substrate or binding partner present. A deep understanding of protein dynamics is therefore critical for understanding protein function, and solution-state NMR spectroscopy is one of the most accurate and precise experimental methods for this characterization (Figure 1D). Proteins undergo a wide range of motions across a varying timescales,[15,17,24–31] which can broadly be classified into two categories: (1) fast motion (ps to ns) characteristic of local backbone and side-chain fluctuations and (2) slow motion (μs, ms, or s) typically characteristic of larger conformational exchanges. Over the past four decades, a battery of experimental approaches has been developed to quantify dynamics across these timescales (Figures 1C–1E). Fast motions can be analyzed by measuring longitudinal spin-relaxation rates (R_1), transverse spin-relaxation rates (R_2), and steady-state heteronuclear NOEs on proton-bound ^{13}C or ^{15}N nuclei.[17] Heteronuclear {^1H}-^{15}N NOEs are particularly sensitive to ps motions, while longitudinal relaxation rates (R_1) are sensitive to motions on the ps to ns timescale. Transverse relaxation rates (R_2) are

more sensitive to ns motions, but they also report on chemical and conformational exchange contributions from slower μs to ms dynamics. Hence, dynamics on the μs to ms timescale can be characterized in terms of an exchange contribution to R_2, called R_{ex},[32] such that the observed transverse relaxation rate is given by, $R_2^{obs} = R_2^0 + R_{ex}$, where R_2^0 represents the intrinsic rate of transverse relaxation rate of the nucleus typically originating from dipole-dipole and chemical shift anisotropy interactions. The values of R_1, R_2, and the NOE can be determined using well established methods,[17,33] and further analysis can be performed using the model-free formalism approach.[34,35] Key parameters obtained from a model-free analysis[36–38] include the rigidity of the bond vector (the order parameter, S^2), the timescales of the local motions (τ_e) the overall molecular tumbling (the rotational correlation time, τ_c), as well as the chemical exchange contributions (R_{ex}). While the model-free formalism can account for R_{ex}, it only provides a qualitative view of μs to ms timescale motion.

Over the past two decades significant advances have been made in NMR techniques for studying conformational exchange and protein dynamics across a wide range of timescales. In the late 1990s, two key experiments were introduced: $R_{1\rho}$ (for k_{ex} between 4,000 s^{-1} and 1,000 s^{-1})[18,19] and CPMG-based relaxation dispersion experiments (for k_{ex} between 4,000 s^{-1} and 200 s^{-1}).[39–43] These experiments effectively suppress the contributions from chemical exchange processes (R_{ex}) in a controlled manner by applying an external radio frequency (RF) field of varying strength. In the $R_{1\rho}$ experiments, R_{ex} suppression is achieved by "spin-locking" the nucleus of interest with a spin-lock field that is either on-resonance or off-resonance. On the other hand, in CPMG-based relaxation dispersion experiments, a

series of refocusing π pulses are applied at regular intervals during a relaxation delay, which effectively refocuses the effects of the chemical exchange processes, including R_{ex} contributions. By varying the intensity and duration of the spin-lock field in $R_{1\rho}$ experiments and the inter-pulse delays in the CPMG-based relaxation dispersion experiments, exchange dynamics occurring on the ms timescale can be characterized. These techniques have been widely used to investigate protein dynamics within the μs to ms range and gain insights into the structure and dynamics of low-populated excited states of proteins sampled at the ms timescale.[44–47] The analysis of relaxation dispersion profiles acquired at multiple magnetic field strengths provides information about exchange rates, populations, and chemical shifts of the excited state(s).

In the early 2000s and 2010s, the CEST[43,48–50] and DEST[51] experiments were adapted to study slower timescale exchange processes within a window of 40 s^{-1} < k_{ex} < 500 s^{-1}. By effectively and selectively saturating protein resonances in the low-populated state with a low-power RF field and monitoring the exchange of saturation to the ground state, one can obtain information about ms dynamics and sparsely populated states. CEST experiments, like CPMG-based relaxation dispersion experiments, offer insight into the populations of sparsely populated exited states, exchange rates, and chemical shifts. Initially, both ^{15}N and ^{13}C versions of these experiments were developed for characterizing ms dynamics in small (<10 kDa) to medium size (<30 kDa) proteins, at both backbone[46] and side-chain positions.[52] Subsequently, with the introduction of the methyl-TROSY experiment in 2003 for studying large proteins (up to 1 MDa), methyl-TROSY-based pulse sequences[53] for the CPMG-based relaxation dispersion and CEST experiments were developed.[41,54,55] These experiments have been instrumental in characterizing ms dynamics in large proteins and macromolecular complexes, thereby complementing cryo-EM studies.[43,56] In recent years, further advancements have been made in NMR techniques for studying conformational exchange including developments for functional side-chains.[57,58] These include a multiple-quantum CEST method focusing on the side-chain guanidinium group of the arginine amino acid and amine of lysine side chains[57,59,60] and diverse sets of methods to characterize lysine side-chains and carboxylic acids.[61] These methods provide insights into conformational dynamics, ligand binding, and other interactions involving charged residues in proteins.

Overall, the application of NMR techniques aimed at characterizing conformational dynamics has significantly advanced our understanding of protein function. By characterizing dynamics and conformational exchange, these methods have enabled researchers to explore the complex behavior of proteins across various timescales, thereby providing valuable insights into structure-function relationships.

Identification and characterization of allosteric/regulatory sites in proteins

Although AlphaFold2 and ESM-2 have provided significant advancements in protein structure prediction (Figure 1A), these predictions are often insufficient for identifying regulatory sites involved in allosteric regulation (Figure 1E). Allosteric regulation remains a prevalent and effective mechanism for controlling enzymatic and binding activity by connecting distal sites within a protein, for example the active and periphery sites. Therefore, the identification and characterization of allosteric sites in proteins plays an important role in understanding protein function and developing specific allosteric inhibitors and activators.

A fundamental question in the study of allostery revolves around the mechanism by which changes in one region of a protein propagate to other regions, thereby regulating binding or enzymatic activity. Numerous examples, such as investigations of the PDZ domain,[62] catabolite activator protein (CAP),[63] adenylate kinase,[64] calmodulin,[65] heat shock protein 90,[66,67] and human histone deacetylase 8 (HDAC8),[31] provide compelling evidence that allostery can induce perturbations to the internal dynamics of a protein or enzyme. These perturbations, in turn, can induce conformational transitions and alter the dynamics of the active site. Conversely, perturbations at the active site can also influence the dynamics of the allosteric sites in certain cases. Allostery, being closely linked to changes in structure and dynamics resulting from ligand binding, can be studied very effectively by NMR spectroscopy. NMR uniquely allows one to probe multiple conformational states simultaneously, providing an advantage for identifying and characterizing allosteric sites and their impact on enzymatic activity. In the following, we provide a summary of NMR-based tools and methods that can be applied for this purpose.

Chemical shift perturbation (CSP) analyses[68] using ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra (for proteins <30 kDa) and methyl-TROSY spectra (for proteins up to 1 MDa) can be employed to identify allosteric regulatory regions within a protein. This powerful method involves comparing spectra of the free and, for example, ligand-bound species, enabling the detection of significant CSPs in distal residues, thus suggesting the presence of potential allosteric sites. Similarly, this method can be applied to compare spectra with and without known potential allosteric effectors to identify allosteric communication pathways.[67,69] The identification of allosteric sites can also be facilitated through the analysis of PREs by identifying the protein residues affected by a PRE label.[70–72] Furthermore, since allosteric communication is often a dynamical process, relaxation measurements, such as CEST and CPMG (introduced previously), provide valuable information on conformational dynamics and the flexibility of regions involved in allosteric regulation.[73–75] For example, CSP analyses in conjunction with methyl-TROSY NMR experiments were applied to probe changes in methyl groups around allosteric sites in the WD40 domain of Cell division control protein 4 (Cdc4) upon titration with its partner, Sic1-derived multi-CPD substrate (pSic1).[76] Similarly, we have recently applied methyl-TROSY NMR and multiple-quantum methyl-TROSY CPMG relaxation dispersion experiments to investigate the structural and dynamical coupling throughout the HDAC8 enzyme and identified a regulatory site 21 Å away from the enzyme's active site.[31] Furthermore, we also observed bidirectional communication between the active and regulatory sites via both titration of a competitive inhibitor and site-directed mutagenesis.[31]

It is important to note that functional studies, such as biochemical assays of the wild type protein and the protein with mutations introduced near the potential allosteric site, are crucial for confirming allosteric effects and determining the

Structure
Review

CellPress
OPEN ACCESS

functional consequences of the identified site. These studies aid in understanding the impact of allosteric sites on ligand binding and catalytic activity.[31,77,78] Ultimately, by correlating the NMR-derived structural and dynamical information with functional consequences, the significance of the allosteric sites in protein regulation can be effectively established. This knowledge can inform the design of targeted drugs that modulate protein function by targeting allosteric sites, thereby guiding the development of therapeutic interventions.

NMR complements cryo-EM in the study of macromolecular complexes

Cryo-EM has gained widespread popularity in the last decade as a valuable method for analyzing the structure of large proteins and macromolecular complexes. However, obtaining atomic-level details, especially for flexible regions within these molecules, remains a challenge for many systems. Currently, most cryo-EM structures have resolutions of 5–6 Å. Nevertheless, it is now increasingly common for cryo-EM structures to have resolutions of 3–4 Å, with some structures below 3 Å.[79,80] To gain detailed information about specific regions or domains within the complex, NMR is often instrumental. Methyl-TROSY based NMR spectroscopy enables the investigation of conformational dynamics and interactions within macromolecular complexes. In certain cases, a combination of NMR and cryo-EM data have been used to obtain a more comprehensive understanding of the structure and function of macromolecular complexes.[56,81]

Recent advances have led to an integrated use of NMR and cryo-EM for the determination of large complex structures with high precision and accuracy. For instance, an integrated approach was employed to determine the structure of the 468 kDa dodecameric aminopeptidase tetrahedral aminopeptidase (TET2) with a resolution below 1 Å.[82] Moreover, molecular dynamics (MD) simulations, in conjunction with solution-state chemical shifts and solid-state magic angle spinning (MAS) NMR, were used to refine the cryo-EM structure.[80] This demonstrates the potential of NMR spectroscopy to complement cryo-EM and computational methods to improve the resolution of complex macromolecular structures.

G-protein coupled receptors (GPCRs) are cell membrane receptors that play a crucial role in signal transduction and impact various physiological processes. They are frequently targeted by pharmaceutical drugs, with over 30% of FDA-approved drugs currently targeting GPCRs. While X-ray crystal structures of various monomeric GPCRs have been deposited in the Protein DataBank, there are ongoing debates as to whether GPCRs form higher order complexes.[83,84] Solution-state NMR has been utilized to investigate the activation and regulation of GPCRs. However, studying GPCRs using NMR presents unique challenges, particularly in regard to protein labeling. Typical protein labeling methods employing *E. coli* are not applicable due to the post-translational modifications that GPCRs undergo. Consequently, eukaryotic expression is necessary for GPCR production. Additionally, to take full advantage of high resolution in ^{15}N, ^{1}H transverse relaxation-optimized spectroscopy (TROSY) and methyl-TROSY based NMR experiments, deuteration of GPCRs is required. Unfortunately, eukaryotic organisms such as insect and mammalian cells do not tolerate D_2O well; however, successful expression of deuterated human GPCRs

was achieved using *P. pastoris* in D_2O media.[85,86] Scientists are now employing two different strategies, ^{19}F NMR and methyl labeling based NMR,[87–90] to enhance the understanding of GPCR activation and regulation, as well as to expand GPCR-targeted drug design. Recent NMR investigations using ^{19}F and methyl probes have shed light on the dynamical characteristics of various receptor states, revealing the pathways and intermediate stages involved in their activation.[88,89,91,92] This dynamical perspective goes beyond the static perspective of GPCRs obtained through alternative techniques, enriching our understanding and ability to target of GPCRs.

BIOMOLECULAR NMR TO CHARACTERIZE FLEXIBLE PROTEINS AND FLEXIBLE REGIONS OF PROTEINS

Introduction to IDPs and IDRs

AlphaFold2,[6] ESM-2,[7] and other AI-based protein structure predictors have played pivotal roles in highlighting the strong prevalence of intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs). In contrast to structured proteins whose functions can to some extent be understood with a single conformation alone (which now can be reliably predicted), IDPs and IDRs lack well-defined three-dimensional structures and instead exist as equilibria of rapidly interconverting states. It is now well accepted that IDPs and IDRs constitute an estimated 30–50% of the human proteome,[93,94] where they often play key biological roles in regulating signaling pathways, facilitating molecular recognition, and participating in the formation of protein complexes.

Although low-confidence scores of AI-based protein structural algorithms correlate well with the presence of disordered regions,[95] these tools cannot reliably produce ensembles of conformational states that these regions adopt, and they sometimes fail to predict how such distributions change in the presence of binding partners.[96] Other computational techniques, including all-atom MD simulations and ensemble generators, offer a versatile toolset for modeling diverse conformational ensembles adopted by IDPs and IDRs across a wide range of timescales,[97–99] spanning from the ns to the μs and even ms regimes. In certain cases, simplifying the system with coarse-grained models, where atoms are represented as "beads," enables the exploration of even longer timescales.[98] Recently, such coarse-graining approaches have paved the way for large-scale simulations of entire proteomes.[100,101] Leveraging the vast amounts of data generated by these simulations, coupled with the power of deep learning, it was possible predict general properties of disordered proteins, such as radii of gyration from sequence alone.[100,101] Nevertheless, both all-atom and coarse-grained MD suffer from sampling limitations and force field inaccuracies, such that different parameter sets predict dramatically different ensembles for the same disordered sequence, highlighting the importance of experimental approaches, especially if atomic insight is desired.

Experimental structural techniques such as X-ray crystallography and cryo-EM generally fail to provide atomic information about highly dynamic protein regions, like those found in IDPs and IDRs. Consequently, the Protein Data Bank, on which AI-based protein structure predictors are trained, are generally devoid of detailed structural information on these highly

prevalent biomolecules. NMR spectroscopy, however, is uniquely suited to characterize the behavior and interactions of these otherwise elusive proteins at the atomistic level on multiple timescales. Below we briefly outline NMR-based methods used to probe the structures, dynamics, and interactions of IDPs and IDRs, and highlight applications of NMR to characterize the many diverse interaction types of these prevalent biomolecules.

NMR methods to probe the structures, dynamics, and interactions of IDPs and IDRs

Conformational ensembles of IDPs and IDRs can be probed with several NMR techniques including chemical shifts, scalar couplings, and residual dipolar couplings (RDCs) (Figure 1C).[97,102,103] These NMR measurements provide valuable insights into the equilibrium states of IDPs and IDRs, and serve as crucial tools to validate, restrain, or reweight structural ensembles obtained from computational approaches.[97,102,104–107] For example, the 2D ^{1}H, ^{15}N HSQC experiment provides data on each ^{1}H-^{15}N spin-pair in the peptide backbone for non-proline residues and amide-containing side chains. This experiment provides a "fingerprint" of the protein, and immediately reveals whether a protein is disordered or folded. Disordered proteins typically display spectra with limited chemical-shift dispersions in the ^{1}H$_N$ dimension, indicating that each residue experiences a similar, solvent-exposed environment, whereas residues in folded proteins experience a wide range of chemical environments, including those that are solvent-exposed or buried in hydrophobic cores, resulting in a large dispersion of corresponding signals.[108] This approach was used to elegantly demonstrate that the regions of the disordered eukaryotic translation initiation factor 4E binding protein 2 folds into a well-defined structure upon phosphorylation.[109] Additionally, an observable called the "secondary chemical shift"[110,111] can be calculated for several nuclei within each amino acid residue including ^{1}H$_\alpha$, ^{1}H$_N$, ^{15}N, ^{13}CO, ^{13}C$_\alpha$, and ^{13}C$_\beta$. By comparing observed secondary chemical shifts to reference values for random coils, one can quantify secondary structural propensities and localize regions of disorder within a given sequence.[102,110,111]

Due to their highly extended states, disordered proteins have a large degree of solvent exposure and undergo substantial hydrogen exchange with the bulk water, particularly at physiological pH and temperatures (pH > 7.0 and T > 25°C). Signals from labile hydrogens can thereby experience extreme line broadening and loss of intensity. Measurements under physiological conditions (pH 7.4, 37°C) can be acquired using ^{13}C-direct detection, such as the ^{13}CO–^{15}N (CON) experiment, an alternative to the ^{1}H-^{15}N HSQC, which is insensitive to hydrogen exchange.[112,113]

While conformational and solvent exchange can contribute to signal loss, these phenomena can be exploited to garner further information about the dynamics of IDPs on the ps to ms timescale using spin-relaxation measurements, hydrogen-deuterium exchange, and PREs. Relaxation measurements, including CPMG[114] and CEST,[115,116] provide insights into the molecular motions on the ms timescale of the different regions within IDPs, shedding light on the flexibility and conformational dynamics of these proteins. Hydrogen-deuterium exchange experiments allow for the characterization of solvent accessibility, providing information about regions that are protected or exposed within the IDP. PREs can be employed to investigate the spatial proximity of IDPs to paramagnetic species, enabling the identification of transient or weak long-range interactions that may be crucial for the function of IDPs.[117] Incorporating data that report on dynamical properties and interactions[118,119] with equilibrium structural data further enhances our understanding of IDP behavior and functionality. The interplay between structure and dynamics is essential for the diverse functional roles of IDPs and IDRs, such as their involvement in protein-protein interactions, signaling pathways, and molecular recognition processes.

Application of NMR to probe IDP/scaffold interactions

IDPs and IDRs have exceptionally diverse interaction modes, including folding-upon-binding,[120] dynamic (or "fuzzy") complex formation,[121] and fold-switching mediated by post-translational modifications. The affinities of complex formation involving IDPs and IDRs also span an enormous range, from millimolar[122] to picomolar.[123] NMR has made significant contributions in elucidating and precisely localizing these interactions, enabling scientists to determine whether conformations fold prior to interaction, local structural changes are induced upon interaction, or if folding even occurs at all upon interaction.[124]

Folding-upon-binding of the disordered domain of Sendai virus nucleoprotein was characterized using NMR to understand how this domain interacts with the folded C-terminal domain of the phosphoprotein. A combined analysis of ^{1}H$_N$, ^{13}CO, and ^{15}N CPMG experiments was employed to characterize a three-state, dynamic interaction between the two partners. A "conformational funnelling" mechanism was observed, in which the disordered domain first adopts a helical conformation, which in turn is stabilized non-specifically on the surface of its folded partner. This helix then localizes into a specific binding site at a rate consistent with the dynamics of the folded domain.[114] In another recent study, multi-nuclear CEST experiments were employed to demonstrate that the disordered DNA binding domain of the cytidine repressor (CytR) transiently populates a folded excited state, responsible for binding DNA via a "conformational selection" mechanism. The structure of this excited state was elucidated using chemical shifts and RDCs.[116]

DEST is another NMR method that allows one to characterize the exchange between NMR-"visible" and NMR-"invisible" species, particularly complexes larger than 1 MDa or paramagnetically broadened states.[51,125] In this method, the highly elevated R_2 values of the large species allow for partial saturation by a weak RF field, where the magnetization of the NMR-"visible" species is unaffected. Through chemical exchange, the partial saturation is transferred to the NMR-"visible" species and recorded as a decrease in signal intensity of the NMR-"visible" species. The decrease in signal intensity as a function of the RF field offset is measured, and the resulting shape of the profiles reflect residue-specific parameters of the NMR-"visible" species interacting with the otherwise invisible state. This approach was applied to probe how the dynamic 40- and 42-residue disordered amyloid-β peptides interact with aggregated species, associated with Alzheimer's disease, at the residue-specific level.[126] It was observed that the first eight residues of amyloid-β do not directly interact with the aggregated species, but instead exist in a mobile "tethered" state, whereas

Structure
Review

the hydrophobic central core of the peptides and hydrophobic regions of the C-termini are in direct contact with the aggregated, fibrillar species. It was also observed that the R_2 values were significantly larger for C-terminal residues of the 42-residue variant than those of the 40-residue variant, potentially explaining the dramatic difference in aggregation rate.

NMR has also been instrumental alongside single molecule Förster resonance energy transfer (FRET) experiments and coarse-grained simulations to characterize a picomolar complex between disordered and positively charged histone H1 and the disordered negatively charged nuclear chaperone prothymosin-α. Chemical shift perturbations, signal intensity changes, and ^{15}N R_1 and R_2 measurements were employed to localize a binding region of approximately 60 residues. The analysis demonstrated that the complex was extremely dynamic, driven by non-specific charge complementarity, and that no structural changes occur upon complex formation.[123]

Application of NMR to probe IDPs and IDRs undergoing LLPS

Liquid-liquid phase separation (LLPS) is a process in which certain biomolecules, including proteins and nucleic acids, form distinct liquid-like compartments in the absence of membranes within cells at physiological concentrations and conditions. Proteins that undergo LLPS often contain IDRs with low complexity sequences and remain dynamic in solution.[127] Sometimes referred to as biomolecular condensates or membraneless organelles, these compartments play crucial roles in cellular organization, signal transduction, and gene regulation. Solution-state NMR offers unique advantages for probing the structural and dynamical properties of IDPs inside liquid-like assemblies[128] given the ability to analyze biomolecules at the atomistic level in their native liquid state in conditions that either mimic cellular environments or even directly in-cell.

LLPS is often described in the context of the two phases: (1) the dilute phase, in which the biomolecule of interest is at a concentration below that required for phase separation, and (2) the coacervate (or condensed) phase, in which the biomolecule of interest is at or above a concentration required for phase separation. There are a variety of techniques to prepare samples containing only the dilute or condensed phase,[128] however, some samples are biphasic (coacervates suspended in the dilute phase). The resulting NMR spectra contain significant overlapping signals from biomolecules in both environments. In these cases, relaxation and diffusion editing can be used to isolate signals from the dispersed or condensed phase, respectively. For example, HSQC measurements of an elastin-like polypeptide enriched in hydrophobic residues undergoing LLPS yields poorly resolved peaks.[129] However, by using an R_2 relaxation filter to remove signals with R_2 relaxation rates above 5 s^{-1} (thus selecting for fast tumbling), Sharpe et al. were able to select for the dilute phase, as biomolecules in the condensed phase have high R_2 rates due to the high viscosity of the environment.[129] Conversely, in the same study, diffusion editing was employed to remove signals arising from biomolecules with a diffusion rate faster than 10^{-7} cm^2 s^{-1}, thus selecting for biomolecules from the coacervate phase.[129]

Whether measurements are taken directly in dilute phases, condensed phases, or filtered in biphasic samples, chemical shifts have played a crucial role in demonstrating that IDPs with low complexity domains remain disordered in the condensed phase via the persistence of low peak dispersion.[129–131] Perturbations of chemical shifts and signal intensities can also be used to map binding and self-interactions. For example, chemical shift measurements acquired using the CON experiment identified residue-specific patterns mediating the interactions between the C-terminal disordered regions of the translational regulators FMRP and CAPRIN1 that repress translation by deadenylating mRNA.[132]

In addition to chemical shifts, NOEs, coupled with differential isotopic labeling schemes and heteronuclear selection have been used to distinguish between intra- and intermolecular interactions of less than 6 Å in LLPS.[128,129,131,133] Furthermore, intra- and intermolecular PRE measurements have been used to monitor disruption of long-range contacts via phosphorylation of the low complexity domain of fused in sarcoma (FUS).[134] Dynamics on the ps to ns timescale can also be probed for phase separating systems. For example, the phase-separating low-complexity domain of transactive response DNA-binding protein 43 (TDP-43) was shown to contain short regions of transient helicity of approximately 20-residues in length in its dilute phase, via an increase in R_2 values and the observation of heteronuclear NOEs.[130] Other NMR methods, including off-resonance $R_{1\rho}$ experiments, were used to demonstrate that residues of the germ granule protein Ddx4 exchange with a minor state with a rate of approximately 18 s^{-1} in the condensed phase.[135]

To date, various NMR methods have demonstrated that IDPs remain in their disordered states in the condensed phase, however, the dynamics are often restricted. Due to the complexities of LLPS sample preparation (e.g., limited stability and protein aggregation) minimal models such as single domains of full length proteins are often studied. Advanced techniques such as segmental isotopic labeling and in-cell NMR offer promising avenues for future research, allowing scientists to enhance the understanding of biological systems by increasing their complexity and maximizing their relevance in a biological context.

Application of NMR to probe small-molecule binding to IDPs

NMR has played an important role in elucidating mechanisms of small molecule-binding to IDPs.[104,136,137] Because of their lack of well-defined binding pockets, the mechanisms underpinning small-molecule interactions with disordered proteins are an open area of research. Some small molecules have been reported to have extremely subtle effects on the protein chemical shifts.[104,136,138] We recently demonstrated that ligand-detected ^{19}F transverse relaxation rates (R_2) are highly sensitive to the interaction between a small molecule and the disordered domains of non-structural protein 5A from hepatitis C virus, in contrast to chemical shift perturbations which are minimally sensitive to this interaction. By combining R_2 measurements, chemical shift perturbations, and diffusion ordered spectroscopy (DOSY) measurements, we could calculate the affinity of the interaction to be approximately 300 μM and the rotational correlation time of the bound state (τ_c) to be approximately 50 ps suggesting that the small molecule remains highly dynamic in the bound form. These results suggest that ligand-detected ^{19}F transverse relaxation measurements could represent a highly effective screening

strategy to identify molecules capable of interacting with these traditionally elusive, dynamic biomolecules.[138]

Since their discovery, IDPs and IDRs have challenged paradigms in structural biology. These biomolecules can be functional in their highly dynamic forms, and in fact, it is their extreme flexibility that often enables them to regulate their many diverse functions. Despite significant progress in protein structure prediction, current approaches cannot yet reliably predict atomistic, conformational ensembles of IDPs and IDRs. The primary limitation lies in the scarcity of reliable training data for new AI tools, which impedes the development of accurate models. MD simulations have the potential to address this limitation by providing enormous amounts of relevant data; however, inaccuracies in force fields and sampling limitations hinder their accuracy, resulting in AI models that may reflect these biases. NMR spectroscopy offers a comprehensive toolkit for accurately characterizing these conformationally heterogeneous biomolecules and their interactions across multiple timescales. Integrating MD and NMR approaches, two highly complementary techniques, holds tremendous promise in reliably and precisely characterizing the structural properties of IDPs.[97,102,104,105] With sufficient experimental data and computational resources, input has been[139] and will continue to be integrated into AI-driven protein ensemble predictors and "back-calculators,"[140] enabling a deeper understanding of the intricate relationships between structure, dynamics, and function in these highly dynamic proteins.

DIRECT APPLICATIONS OF DEEP LEARNING WITHIN NMR SPECTROSCOPY

Advances in AI and deep learning such as protein structure predictors indirectly underscore the need for further analyses of protein motions, an area to which NMR can offer unparalleled insight as discussed previously. Within NMR spectroscopy, neural networks have been developed for peak-picking,[141] for classification of peaks,[142,143] and for automated assignment[144] of NMR spectra for more than 30 years. Widely used tools such as TALOS+,[145] that predicts backbone dihedral angles from chemical shifts, and the random coil index (RCI) method,[146] that predicts order parameters from chemical shifts, are also based on neural networks and machine learning. However, over the last half decade there has been a recent reinvigoration in this area; AI and deep learning tools have been recently integrated directly with the transformation and analysis of NMR data to improve the methodology. These tools can transform complex NMR data into easily interpretable spectra and are now starting to assist the NMR user with complex tasks of analyzing NMR spectra more effectively (Figure 2). In the following, we will initially discuss how deep learning tools have been developed to robustly reconstruct sparsely and non-uniformly sampled (NUS) spectra. We will also discuss how deep learning has emerged as a powerful tool for virtual decoupling and enhancing the resolution of NMR spectra, enabling significant advancements in both solution biomolecular NMR[147] and solid-state NMR spectroscopy.[148] Finally, we will discuss new developments of deep learning to perform autonomous analysis of complex NMR data, which includes both the detection of peaks in NMR spectra and the analysis of complex NMR data reporting on the chemical and conformational exchange of proteins. The focus below will mainly be on biomolecular and related applications, although impressive developments have also appeared within other fields of NMR spectroscopy.

Deep learning tools to transform NMR spectra

Unlike traditional tools for analysis and transformation of data, where a pre-defined algorithm is required, in deep learning, deep neural networks (DNNs) are trained to learn a mapping between an input and a desired output (Figure 2). A key advantage of deep learning compared to traditional tools and algorithms is that it has been particularly successful at performing tasks that are often intuitive to humans but difficult to formalize into a pre-defined algorithm.[3,4] Within supervised deep learning, the main technique used within NMR, training the DNNs require a large amount of so-called "training data," where pairs of input and desired output are provided, and sophisticated minimisers, such as ADAM or RMSprop, are employed to optimize the weights within the layers of the DNN to learn the mapping.

Within most fields of NMR spectroscopy, including biomolecular NMR, the Fourier transform is the main function or mapping used to transform raw time-domain data into frequency-domain data, enabling analysis downstream. Such traditional analyses typically involve manual inspections at each stage of the analysis and least-squares fitting is often used to assess the obtained spectral parameters with a pre-defined algorithm, for example, to extract information about structure, dynamics, or kinetics of a system. However, this traditional linear workflow imposes constraints on experimental design and data analysis. Deep learning approaches have demonstrated the potential to overcome these constraints and outperform traditional methods in various scientific domains. Therefore, there is significant potential for deep learning approaches to enhance transformation and analysis stages within NMR, improving efficiency, utility, and user-friendliness.

Proof-of-concept studies initially focused on the reconstruction of non-uniform-sampling (NUS) biomolecular NMR data.[149–152] Sparse sampling and NUS techniques are commonly employed in biomolecular NMR to acquire high-dimensional spectra,[153–155] such as 4D methyl-methyl nuclear Overhauser effect spectroscopy (NOESY) spectra[156] or spectra for chemical shift assignments of IDPs.[157] These approaches significantly expedite data acquisition while preserving the quality of spectral information by strategically selecting only a subset of data points for collection. The reduction in experimental time achieved by sparse sampling and NUS techniques are particularly important for biomolecules with limited stabilities. For the deep learning developments in NUS NMR, the input data consisted of sparsely sampled NMR spectra and the output was the fully sampled spectrum, either in the time domain[149,150] or in the frequency domain.[151,152] For reconstructions in the time domain, the DNN effectively predicts the missing data points, whereas for reconstructions in the frequency domain the DNN removes artifacts associated with sparse and non-Nyquist sampling. Overall using a DNN for reconstruction of sparsely sampled data were demonstrated by several groups to be effective with high accuracy, even when the sampling was very sparse. Although DNN reconstructions typically did not substantially outperform traditional algorithms used in biomolecular NMR, such as multidimensional decomposition (MDD) NMR,[158] istHMS,[159] and sparse multidimensional iterative

Structure
Review

lineshape-enhanced (SMILE) reconstruction,[160] there are several advantages in using DNNs, including faster reconstructions[151,152] and ease of implementation into automated pipelines. Robust DNN network architectures, such as FID-Net,[150] which are easy and fast to train, are now available to generally transform biomolecular NMR spectra.

One notable finding from these initial developments is that the DNNs can be trained on fully synthetic data,[149,150] which differs from many other machine learning fields where obtaining, curating, and annotating training data are often a bottleneck. Additionally, the well-understood theory for NMR spectroscopy allows for accurate simulation of nearly all possible NMR experiments either using simple home-written or specialized software, such as SPINACH,[161] or SIMPSON.[162] This establishes a solid foundation for developing and utilizing supervised deep learning methods for the analysis and transformation of complex NMR data.

Deep learning and virtual homonuclear decoupling
Homonuclear scalar couplings play an important role in all areas of NMR, where they are used to transfer magnetization between various sites in both small molecules and in macromolecules in solution. The early development of multi-dimensional NMR experiments, such as COrrelated SpectroscopY (COSY) and TOtal Correlation SpectroscopY (TOCSY) experiments, hinge on transfer between spins mediated by homonuclear scalar couplings. Even today, most chemical shift assignments for small molecules and proteins in solution heavily depend on transfers via these couplings. However, the presence of homonuclear couplings between neighboring nuclei can adversely impact the quality of NMR spectra, leading to overlapping peaks that hinder the identification and quantification of signals from individual nuclei. Effectively, these scalar couplings double, triple, or quadruple the number of peaks in the observed NMR spectrum, significantly reducing the resolution and often complicates interpretation.

To remove homonuclear couplings and sharpen peaks, decoupling schemes are commonly employed. However, these schemes can cause a range of deleterious effects including sidebands, Bloch-Siegert shifts, and severe relaxation losses.[163,164] To address this, the preferred method for eliminating homonuclear couplings in biomolecular NMR spectra involves virtual decoupling techniques such as in-phase/anti-phase (IPAP)[165] and (double in-phase/anti-phase) DIPAP.[166,167] These methods require recording several sub-spectra, and the final decoupled spectrum is obtained by taking linear combinations of these sub-spectra. Although these approaches have proven useful, they do come with several drawbacks. Several spectra must be recorded, and additional delays and frequency-selective pulses are often required to invert the passive spin(s) for virtual decoupling, all of which lead to reduced sensitivity and, in some cases, severe artifacts in the final spectrum.

In recent years, deep learning has emerged as a powerful tool for virtually decoupling of NMR spectra. In these applications, the DNN is trained on pairs of spectra, where the input spectra contain scalar couplings, while the target output spectrum is the decoupled spectrum. Similar to the reconstruction of sparsely sampled spectra, the training data can be synthetically generated using scenario-specific parameters. Our recent work on virtual

decoupling of $^{13}C_\alpha$-detected spectra has demonstrated the accurate training of DNNs to virtually decouple different multiplet structures in a single pass.[147] $^{13}C_\alpha$-detected NMR spectra have previously been shown to provide valuable information on IDPs for assignments,[167,168] however, these spectra lead to a complex doublet-of-doublet pattern originating from scalar couplings between $^{13}C_\alpha$ and ^{13}CO as well as $^{13}C_\alpha$ and $^{13}C_\beta$ (for non-glycine residues). Methods such as DIPAP or double spin-state selective excitation (DS_3E) methods have previously been used to virtually decouple these experiments,[166] although these decoupling techniques require four sub-spectra to be recorded and often lead to artifacts in the resulting spectra for serine, threonine, and glycine residues. In contrast, using DNNs, we showed that $^{13}C_\alpha$-detected spectra can be robustly decoupled, that only one input spectrum is required, and that the DNN virtually decouples all residue types, including, serine, threonine and glycine.[147] More importantly, virtual decoupling with DNNs eliminates the need for implementing frequency-selective pulses into the pulse sequences, allowing for enhanced signal-to-noise ratios and accurate decoupling of all sites, irrespective of the chemical shift of the passive spin. Furthermore, deep learning-based tools have also been developed to assist in obtaining decoupled NMR spectra of small molecules. Recent work[169] showcases the utilization of deep learning in the context of pure shift NMR,[170] a technique that enhances the resolution of NMR spectra, typically of organic compounds, with overlapping signals. The authors first obtain Zangger-Sterk decoupled spectra,[171] which are then processed with a trained DNN to generate high-resolution spectra devoid of couplings. This approach can effectively separate overlapping peaks, enabling a more accurate determination of chemical shifts. This methodology was successfully applied to various systems, such as natural products and synthetic molecules, facilitating improved spectral analysis.

The integration of deep learning techniques within NMR spectroscopy has revolutionized the field by enabling the transformation of complex spectra of both small molecules and biomolecules into high-resolution spectra. These advancements provide enhanced resolution, sensitivity, and interpretability, empowering researchers to gain deeper insights into the structures and dynamics of the molecules under investigation.

Deep learning tools for the analysis of NMR data
The Achilles' heel of NMR spectroscopy is often the extraction of the relevant parameters from complex data to report on the properties of the molecule or system under investigation. This includes tasks such as chemical shift assignment, interpretations of cross-peaks in NOESY spectra to derive inter-atomic distances, and the analysis of CPMG[42] and CEST[172] data to gain insight on functional dynamics and transiently populated states.[173]

Although advancements in isotope labeling schemes[174] and NMR-based methods have enabled a plethora of experiments to investigate diverse biochemical properties, these techniques often require laborious manual or semi-manual procedures and often require the knowledge of specialists with decades of experience. Consequently, sophisticated NMR methods that offer invaluable insights into biological systems are effectively only accessible to a limited fraction of scientists in biochemistry and structural biology. However, this scenario is poised to change with the emergence of autonomous and robust AI and

deep learning tools, which will make even the most sophisticated NMR methods available to all scientists.

Automated detection of cross-peaks, called peak-picking, in biomolecular NMR has been an active field of research for decades. The task is very similar to what is often known as segmentation and object detection in the field of AI, and it is therefore not surprising that one of the first analysis tasks in biomolecular NMR, in which DNNs were used to aid the analysis, was peak-picking.[175] Subsequent DNN tools have been developed for picking peaks in complex 2D NMR spectra,[176] although manual inspection and intervention are still often required.

Analysis of CEST spectra can be time-consuming and labour-intensive, even with sophisticated software available for least-squares fitting. Challenges persist in identifying appropriate starting parameters for the fitting procedure and evaluating the outcomes of the analysis. Recently we showed that DNNs can be trained to autonomously analyze CEST data.[177] Notably, even one of the most complicated CEST experiments, the ^1H_N CEST experiment,[55] which exhibits anti-phase CEST "dips," can be analyzed autonomously by DNNs. Specifically, for the ^1H_N CEST experiment, two separate DNNs were trained: one for decoupling the anti-phase ^1H_N CEST profile into an in-phase CEST profile and another for determining the chemical shifts of the exchanging species along with uncertainties. When these two DNNs are combined, they provide a single autonomous tool. Cross-validations on both synthetic and experimental data have demonstrated the accurate determination of chemical shifts and their uncertainties using trained DNNs. The advantages of employing DNNs for analysis of experimental CEST data are many-fold; the analysis becomes autonomous, eliminating the need for end-users to optimize processing parameters thus enabling any scientist interested in chemical and conformational exchange to use these experiments. Furthermore, the autonomous analysis can easily be implemented into automated pipelines, as exemplified recently by the NMROnline—ELECTRO cloud-based analysis software.[178]

Overall, we believe that there is a bright future in integrating deep learning and AI directly within various aspects of NMR spectroscopy. This includes the transformation of NMR spectra to achieve improved resolution and sensitivity, as well as the analysis of NMR data through the generation of automated pipelines using AI tools. The direct implementation of AI offers enhanced accuracy and reliability of NMR measurements while reducing the time and costs associated with performing and analyzing NMR experiments. Particularly intriguing is the prospect of developing AI and NMR as a unified tool, where AI-driven analysis and potential transformations of NMR data are considered in the development of NMR methods. Early indications of this potential have already emerged,[147] as evidenced by simplified pulse sequences that substantially improve sensitivity and resolution through analysis with DNNs rather than classical 1822 Fourier transforms.[179]

CONCLUSIONS AND OUTLOOK

Biomolecular NMR has undergone significant transformations in recent decades, evolving from its traditional role in solving structures of small-to-medium sized proteins[180] to an unparalleled experimental method that provides insights on functional macromolecular dynamics and transient interactions. As such, NMR has become a key complement to experimental structural techniques such as X-ray crystallography, cryo-EM, and AI-based structure prediction,[181] offering orthogonal information about molecular motion and alternative states. Additionally, biomolecular NMR spectroscopy serves as a valuable tool in conjunction with MD simulations. While simulations provide atomistic (or near-atomistic) details of large macromolecular complexes or IDPs, the incorporation of experimental NMR parameters as restraints or cross-validation data helps overcome force field inaccuracies. Finally, in the era of computational structural biology, provided by significant advancements such as AlphaFold2[6] or ESM-2,[7] NMR spectroscopy has become more important than ever. The computational approaches of today excel at generating high-quality structures of large (non-disordered) macromolecular complexes, while biomolecular NMR provides invaluable insights into the flexibility, dynamics, and kinetics of functional structural transitions surrounding these average structures and is especially well-suited for proteins containing IDRs.

It is conceivable that in the near future, AI agents will autonomously operate NMR machines, akin to how self-driving cars navigate with minimal input from passengers. This development will unlock the full potential of NMR spectroscopy, removing the barrier of decades of training that currently limit the technique's accessibility. Combined with data from computational structural biology, including AI-generated structures and potentially AI-analyzed cryo-EM data, any researcher will have the opportunity to explore new ideas and gain a deeper understanding of complex biological systems.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT to proof-read part of the text of this manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

1. Greener, J.G., Kandathil, S.M., Moffat, L., and Jones, D.T. (2022). A guide to machine learning for biologists. Nat. Rev. Mol. Cell Biol. 23, 40–55.

2. Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface 15, 20170387.

3. Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT press).

4. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature 521, 436–444.

Structure
Review

CellPress
OPEN ACCESS

5. Becker, E.D. (1993). A brief history of nuclear magnetic resonance. Anal. Chem. *65*, 295A–302A.

6. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

7. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123–1130.

8. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596.

9. Chatterjee, A., Walters, R., Shafi, Z., Ahmed, O.S., Sebek, M., Gysi, D., Yu, R., Eliassi-Rad, T., Barabási, A.L., and Menichetti, G. (2023). Improving the generalizability of protein-ligand binding predictions with AI-Bind. Nat. Commun. *14*, 1989.

10. Krapp, L.F., Abriata, L.A., Cortés Rodriguez, F., and Dal Peraro, M. (2023). PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. Nat. Commun. *14*, 2175.

11. Ren, F., Ding, X., Zheng, M., Korzinkin, M., Cai, X., Zhu, W., Mantsyzov, A., Aliper, A., Aladinskiy, V., Cao, Z., et al. (2023). AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem. Sci. *14*, 1443–1452.

12. Bertini, I., Donaire, A., Luchinat, C., and Rosato, A. (1997). Paramagnetic relaxation as a tool for solution structure determination: Clostridium pasteurianum ferredoxin as an example. Proteins *29*, 348–358.

13. Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C.M., Poulsen, F.M., and Vendruscolo, M. (2004). Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. J. Am. Chem. Soc. *126*, 3291–3299.

14. Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J. Biomol. NMR *24*, 171–189.

15. Alderson, T.R., and Kay, L.E. (2021). NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. Cell *184*, 577–595.

16. Palmer, A.G., III (2014). Chemical exchange in biomacromolecules: past, present, and future. J. Magn. Reson. *241*, 3–17.

17. Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry *28*, 8972–8979.

18. Akke, M., and Palmer, A.G. (1996). Monitoring macromolecular motions on microsecond to millisecond time scales by R1ρ− R1 constant relaxation time NMR spectroscopy. J. Am. Chem. Soc. *118*, 911–912.

19. Palmer, A.G., and Massi, F. (2006). Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem. Rev. *106*, 1700–1719.

20. Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E. (1994). A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J. Biomol. NMR *4*, 727–734.

21. Koshland, D.E. (1958). Application of a theory of enzyme specificity to protein synthesis. Proc. Natl. Acad. Sci. USA *44*, 98–104.

22. Bennett, W.S., Jr., and Steitz, T.A. (1978). Glucose-induced conformational change in yeast hexokinase. Proc. Natl. Acad. Sci. USA *75*, 4848–4852.

23. Bacon, F., Wyman, J., and Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. J. Mol. Biol. *12*, 88–118.

24. Kovermann, M., Rogne, P., and Wolf-Watz, M. (2016). Protein dynamics and function from solution state NMR spectroscopy. Q. Rev. Biophys. *49*, e6.

25. Sekhar, A., and Kay, L.E. (2019). An NMR view of protein dynamics in health and disease. Annu. Rev. Biophys. *48*, 297–319.

26. Ishima, R., and Torchia, D.A. (2000). Protein dynamics from NMR. Nat. Struct. Biol. *7*, 740–743.

27. Boehr, D.D., Dyson, H.J., and Wright, P.E. (2006). An NMR perspective on enzyme dynamics. Chem. Rev. *106*, 3055–3079.

28. Kay, L.E. (2016). New views of functionally dynamic proteins by solution NMR spectroscopy. J. Mol. Biol. *428*, 323–331.

29. Levy, R.M., Karplus, M., and McCammon, J.A. (1980). Molecular dynamics studies of NMR relaxation in proteins. Biophys. J. *32*, 628–630.

30. Kern, D., Eisenmesser, E.Z., and Wolf-Watz, M. (2005). Enzyme dynamics during catalysis measured by NMR spectroscopy. In Methods in enzymology (Elsevier), pp. 507–524.

31. Werbeck, N.D., Shukla, V.K., Kunze, M.B.A., Yalinca, H., Pritchard, R.B., Siemons, L., Mondal, S., Greenwood, S.O.R., Kirkpatrick, J., Marson, C.M., and Hansen, D.F. (2020). A distal regulatory region of a class I human histone deacetylase. Nat. Commun. *11*, 3841.

32. Palmer, A.G., III, Rance, M., and Wright, P.E. (1991). Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. J. Am. Chem. Soc. *113*, 4371–4380.

33. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry *33*, 5984–6003.

34. Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J. Am. Chem. Soc. *104*, 4546–4559.

35. Lipari, G., and Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. J. Am. Chem. Soc. *104*, 4559–4570.

36. Chen, K., and Tjandra, N. (2008). Extended model free approach to analyze correlation functions of multidomain proteins in the presence of motional coupling. J. Am. Chem. Soc. *130*, 12745–12751.

37. Cole, R., and Loria, J.P. (2003). FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. J. Biomol. NMR *26*, 203–213.

38. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. J. Am. Chem. Soc. *112*, 4989–4991.

39. Loria, J.P., Rance, M., and Palmer, A.G. (1999). A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. J. Am. Chem. Soc. *121*, 2331–2332.

40. Tollinger, M., Skrynnikov, N.R., Mulder, F.A., Forman-Kay, J.D., and Kay, L.E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. J. Am. Chem. Soc. *123*, 11341–11352.

41. Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V., and Kay, L.E. (2004). Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. J. Am. Chem. Soc. *126*, 3964–3973.

42. Sauerwein, A.C., and Hansen, D.F. (2015). Relaxation dispersion NMR spectroscopy. Protein NMR: Modern Techniques and Biomedical Applications, 75–132.

43. Vallurupalli, P., Bouvignies, G., and Kay, L.E. (2012). Studying "invisible" excited protein states in slow exchange with a major state conformation. J. Am. Chem. Soc. *134*, 8148–8161.

44. Hansen, D.F., Vallurupalli, P., Lundström, P., Neudecker, P., and Kay, L.E. (2008). Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: How well can we do? J. Am. Chem. Soc. *130*, 2667–2675.

45. Bouvignies, G., Vallurupalli, P., Hansen, D.F., Correia, B.E., Lange, O., Bah, A., Vernon, R.M., Dahlquist, F.W., Baker, D., and Kay, L.E. (2011). Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. Nature *477*, 111–114.

46. Vallurupalli, P., Hansen, D.F., and Kay, L.E. (2008). Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. Proc. Natl. Acad. Sci. USA *105*, 11766–11771.

47. Hansen, D.F., Vallurupalli, P., and Kay, L.E. (2009). Measurement of methyl group motional parameters of invisible, excited protein states by NMR spectroscopy. J. Am. Chem. Soc. *131*, 12745–12754.

48. van Zijl, P.C.M., and Yadav, N.N. (2011). Chemical exchange saturation transfer (CEST): What is in a name and what isn't? Magn. Reson. Med. *65*, 927–948.

49. Ward, K.M., Aletras, A.H., and Balaban, R.S. (2000). A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). J. Magn. Reson. *143*, 79–87.

50. Zhou, J., and Zijl, P.C.v. (2006). Chemical exchange saturation transfer imaging and spectroscopy. Prog. Nucl. Magn. Reson. Spectrosc. *48*, 109–136.

51. Fawzi, N.L., Ying, J., Torchia, D.A., and Clore, G.M. (2012). Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. Nat. Protoc. *7*, 1523–1533.

52. Bouvignies, G., Vallurupalli, P., and Kay, L.E. (2014). Visualizing side chains of invisible protein conformers by solution NMR. J. Mol. Biol. *426*, 763–774.

53. Yuwen, T., Huang, R., Vallurupalli, P., and Kay, L.E. (2019). A methyl-TROSY-based 1H relaxation dispersion experiment for studies of conformational exchange in high molecular weight proteins. Angew. Chem. *131*, 6316–6320.

54. Rennella, E., Huang, R., Velyvis, A., and Kay, L.E. (2015). 13CHD2–CEST NMR spectroscopy provides an avenue for studies of conformational exchange in high molecular weight proteins. J. Biomol. NMR *63*, 187–199.

55. Yuwen, T., Sekhar, A., and Kay, L.E. (2017). Separating dipolar and chemical exchange magnetization transfer processes in 1H-CEST. Angew. Chem., Int. Ed. Engl. *56*, 6122–6125.

56. Vahidi, S., Ripstein, Z.A., Juravsky, J.B., Rennella, E., Goldberg, A.L., Mittermaier, A.K., Rubinstein, J.L., and Kay, L.E. (2020). An allosteric switch regulates Mycobacterium tuberculosis ClpP1P2 protease function as established by cryo-EM and methyl-TROSY NMR. Proc. Natl. Acad. Sci. USA *117*, 5895–5906.

57. Karunanithy, G., Shukla, V.K., and Hansen, D.F. (2021). Methodological advancements for characterising protein side chains by NMR spectroscopy. Curr. Opin. Struct. Biol. *70*, 61–69.

58. Pritchard, R.B., and Hansen, D.F. (2019). Characterising side chains in large proteins by protonless 13C-detected NMR spectroscopy. Nat. Commun. *10*, 1747.

59. Karunanithy, G., Reinstein, J., and Hansen, D.F. (2020). Multiquantum chemical exchange saturation transfer NMR to quantify symmetrical exchange: Application to rotational dynamics of the guanidinium group in arginine side chains. J. Phys. Chem. Lett. *11*, 5649–5654.

60. Nguyen, D., Chen, C., Pettitt, B.M., and Iwahara, J. (2019). NMR methods for characterizing the basic side chains of proteins: electrostatic interactions, hydrogen bonds, and conformational dynamics. Methods Enzymol. *615*, 285–332.

61. Paquin, R., Ferrage, F., Mulder, F.A.A., Akke, M., and Bodenhausen, G. (2008). Multiple-timescale dynamics of side-chain carboxyl and carbonyl groups in proteins by 13C nuclear spin relaxation. J. Am. Chem. Soc. *130*, 15805–15807.

62. Petit, C.M., Zhang, J., Sapienza, P.J., Fuentes, E.J., and Lee, A.L. (2009). Hidden dynamic allostery in a PDZ domain. Proc. Natl. Acad. Sci. USA *106*, 18249–18254.

63. Popovych, N., Tzeng, S.-R., Tonelli, M., Ebright, R.H., and Kalodimos, C.G. (2009). Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc. Natl. Acad. Sci. USA *106*, 6927–6932.

64. Schrank, T.P., Bolen, D.W., and Hilser, V.J. (2009). Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. Proc. Natl. Acad. Sci. USA *106*, 16984–16989.

65. Igumenova, T.I., Lee, A.L., and Wand, A.J. (2005). Backbone and side chain dynamics of mutant calmodulin−peptide complexes. Biochemistry *44*, 12627–12639.

66. Zhou, C., Zhang, C., Zhu, H., Liu, Z., Su, H., Zhang, X., Chen, T., Zhong, Y., Hu, H., Xiong, M., et al. (2020). Allosteric regulation of Hsp90α's activity by small molecules targeting the middle domain of the chaperone. iScience *23*, 100857.

67. McConnell, J.R., Dyson, H.J., and McAlpine, S.R. (2021). Using NMR to identify binding regions for N and C-terminal Hsp90 inhibitors using Hsp90 domains. RSC Med. Chem. *12*, 410–415.

68. Williamson, M.P. (2013). Using chemical shift perturbation to characterise ligand binding. Prog. Nucl. Magn. Reson. Spectrosc. *73*, 1–16.

69. Velyvis, A., Schachman, H.K., and Kay, L.E. (2009). Application of methyl-TROSY NMR to test allosteric models describing effects of nucleotide binding to aspartate transcarbamoylase. J. Mol. Biol. *387*, 540–547.

70. Otting, G. (2010). Protein NMR using paramagnetic ions. Annu. Rev. Biophys. *39*, 387–405.

71. Bertini, I., Luchinat, C., Parigi, G., and Ravera, E. (2016). NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models (Elsevier).

72. Keizers, P.H.J., and Ubbink, M. (2011). Paramagnetic tagging for protein structure and dynamics analysis. Prog. Nucl. Magn. Reson. Spectrosc. *58*, 88–96.

73. Köhler, C., Carlström, G., Gunnarsson, A., Weininger, U., Tångefjord, S., Ullah, V., Lepistö, M., Karlsson, U., Papavoine, T., Edman, K., and Akke, M. (2020). Dynamic allosteric communication pathway directing differential activation of the glucocorticoid receptor. Sci. Adv. *6*, eabb5277.

74. Grutsch, S., Brüschweiler, S., and Tollinger, M. (2016). NMR methods to study dynamic allostery. PLoS Comput. Biol. *12*, e1004620.

75. Tzeng, S.-R., and Kalodimos, C.G. (2011). Protein dynamics and allostery: an NMR view. Curr. Opin. Struct. Biol. *21*, 62–67.

76. Csizmok, V., Orlicky, S., Cheng, J., Song, J., Bah, A., Delgoshaie, N., Lin, H., Mittag, T., Sicheri, F., Chan, H.S., et al. (2017). An allosteric conduit facilitates dynamic multisite substrate recognition by the SCFCdc4 ubiquitin ligase. Nat. Commun. *8*, 13943.

77. Abdelkarim, H., Hitchinson, B., Banerjee, A., and Gaponenko, V. (2019). Advances in NMR methods to identify allosteric sites and allosteric ligands. Adv. Exp. Med. Biol. *1163*, 171–186.

78. Vahidi, S., Ripstein, Z.A., Bonomi, M., Yuwen, T., Mabanglo, M.F., Juravsky, J.B., Rizzolo, K., Velyvis, A., Houry, W.A., Vendruscolo, M., et al. (2018). Reversible inhibition of the ClpP protease via an N-terminal conformational switch. Proc. Natl. Acad. Sci. USA *115*, E6447–E6456.

79. Geraets, J.A., Pothula, K.R., and Schröder, G.F. (2020). Integrating cryo-EM and NMR data. Curr. Opin. Struct. Biol. *61*, 173–181.

80. Perilla, J.R., Zhao, G., Lu, M., Ning, J., Hou, G., Byeon, I.-J.L., Gronenborn, A.M., Polenova, T., and Zhang, P. (2017). CryoEM structure refinement by integrating NMR chemical shifts with molecular dynamics simulations. J. Phys. Chem. B *121*, 3853–3863.

81. Huang, R., Ripstein, Z.A., Rubinstein, J.L., and Kay, L.E. (2020). Probing cooperativity of N-terminal domain orientations in the p97 molecular machine: Synergy between NMR spectroscopy and cryo-EM. Angew. Chem. *132*, 22609–22612.

82. Gauto, D.F., Estrozi, L.F., Schwieters, C.D., Effantin, G., Macek, P., Sounier, R., Sivertsen, A.C., Schmidt, E., Kerfah, R., Mas, G., et al. (2019). Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. Nat. Commun. *10*, 2697.

Structure
Review

CellPress
OPEN ACCESS

83. Milligan, G., Ward, R.J., and Marsango, S. (2019). GPCR homo-oligomerization. Curr. Opin. Cell Biol. *57*, 40–47.

84. Fuxe, K., Borroto-Escuela, D.O., Marcellino, D., Romero-Fernandez, W., Frankowska, M., Guidolin, D., Filip, M., Ferraro, L., Woods, A.S., Tarakanov, A., et al. (2012). GPCR heteromers and their allosteric receptor-receptor interactions. Curr. Med. Chem. *19*, 356–363.

85. Clark, L.D., Dikiy, I., Chapman, K., Rödström, K.E., Aramini, J., LeVine, M.V., Khelashvili, G., Rasmussen, S.G., Gardner, K.H., and Rosenbaum, D.M. (2017). Ligand modulation of sidechain dynamics in a wild-type human GPCR. Elife *6*, e28505.

86. Eddy, M.T., Lee, M.-Y., Gao, Z.-G., White, K.L., Didenko, T., Horst, R., Audet, M., Stanczak, P., McClary, K.M., Han, G.W., et al. (2018). Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. Cell *172*, 68–80.e12.

87. Pan, B., Liu, D., Yang, L., and Wüthrich, K. (2022). GPCR large-amplitude dynamics by 19F-NMR of aprepitant bound to the neurokinin 1 receptor. Proc. Natl. Acad. Sci. USA *119*, e2122682119.

88. Shiraishi, Y., Kofuku, Y., Ueda, T., Pandey, S., Dwivedi-Agnihotri, H., Shukla, A.K., and Shimada, I. (2021). Biphasic activation of β-arrestin 1 upon interaction with a GPCR revealed by methyl-TROSY NMR. Nat. Commun. *12*, 7158.

89. Manglik, A., Kim, T.H., Masureel, M., Altenbach, C., Yang, Z., Hilger, D., Lerch, M.T., Kobilka, T.S., Thian, F.S., Hubbell, W.L., et al. (2015). Structural insights into the dynamic process of β2-adrenergic receptor signaling. Cell *161*, 1101–1111.

90. Nygaard, R., Zou, Y., Dror, R.O., Mildorf, T.J., Arlow, D.H., Manglik, A., Pan, A.C., Liu, C.W., Fung, J.J., Bokoch, M.P., et al. (2013). The dynamic process of β2-adrenergic receptor activation. Cell *152*, 532–542.

91. Picard, L.-P., and Prosser, R.S. (2021). Advances in the study of GPCRs by 19F NMR. Curr. Opin. Struct. Biol. *69*, 169–176.

92. Frei, J.N., Broadhurst, R.W., Bostock, M.J., Solt, A., Jones, A.J.Y., Gabriel, F., Tandale, A., Shrestha, B., and Nietlispach, D. (2020). Conformational plasticity of ligand-bound and ternary GPCR complexes studied by 19F NMR of the β1-adrenergic receptor. Nat. Commun. *11*, 669.

93. Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. Genome informatics *11*, 161–171.

94. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradović, Z. (2002). Intrinsic disorder and protein function. Biochemistry *41*, 6573–6582.

95. Ruff, K.M., and Pappu, R.V. (2021). AlphaFold and implications for intrinsically disordered proteins. J. Mol. Biol. *433*, 167208.

96. Alderson, T.R., Pritišanac, I., Kolarić, Đ., Moses, A.M., and Forman-Kay, J.D. (2022). Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. bioRxiv *2002*. https://doi.org/10.1101/2022.02.18.481080.

97. Bonomi, M., Heller, G.T., Camilloni, C., and Vendruscolo, M. (2017). Principles of protein structural ensemble determination. Curr. Opin. Struct. Biol. *42*, 106–116.

98. Shea, J.-E., Best, R.B., and Mittal, J. (2021). Physics-based computational and theoretical approaches to intrinsically disordered proteins. Curr. Opin. Struct. Biol. *67*, 219–225.

99. Teixeira, J.M.C., Liu, Z.H., Namini, A., Li, J., Vernon, R.M., Krzeminski, M., Shamandy, A.A., Zhang, O., Haghighatlari, M., Yu, L., et al. (2022). IDPConformerGenerator: A flexible software suite for sampling the conformational space of disordered protein states. J. Phys. Chem. A *126*, 5985–6003.

100. Tesei, G., Trolle, A.I., Jonsson, N., Betz, J., Pesce, F., Johansson, K.E., and Lindorff-Larsen, K. (2023). Conformational ensembles of the human intrinsically disordered proteome: Bridging chain compaction with function and sequence conservation. bioRxiv *2005*, 539815. https://doi.org/10.10101/2023.05.08.539824.

101. Lotthammer, J.M., Ginell, G.M., Griffith, D., Emenecker, R.J., and Holehouse, A.S. (2023). Direct prediction of intrinsically disordered protein conformational properties from sequence. bioRxiv *2005*, 539824. https://doi.org/10.10101/2023.05.08.539824.

102. Heller, G., Yu, L., and Hansen, D. (2022). Characterising Intrinsically Disordered Proteins Using NMR Spectroscopy and MD Simulations. In NMR Spectroscopy for Probing Functional Dynamics at Biological Interfaces (Royal Society of Chemistry), pp. 383–410.

103. Ahmed, R., and Forman-Kay, J.D. (2022). NMR insights into dynamic, multivalent interactions of intrinsically disordered regions: from discrete complexes to condensates. Essays Biochem. *66*, 863–873.

104. Heller, G.T., Aprile, F.A., Michaels, T.C.T., Limbocker, R., Perni, M., Ruggeri, F.S., Mannini, B., Löhr, T., Bonomi, M., Camilloni, C., et al. (2020). Small-molecule sequestration of amyloid-β as a drug discovery strategy for Alzheimer's disease. Sci. Adv. *6*, eabb5924.

105. Heller, G.T., Aprile, F.A., Bonomi, M., Camilloni, C., De Simone, A., and Vendruscolo, M. (2017). Sequence specificity in the entropy-driven binding of a small molecule and a disordered peptide. J. Mol. Biol. *429*, 2772–2779.

106. Rangan, R., Bonomi, M., Heller, G.T., Cesari, A., Bussi, G., and Vendruscolo, M. (2018). Determination of structural ensembles of proteins: restraining vs reweighting. J. Chem. Theor. Comput. *14*, 6632–6641.

107. Bottaro, S., and Lindorff-Larsen, K. (2018). Biophysical experiments and biomolecular simulations: A perfect match? Science *361*, 355–360.

108. Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry *31*, 1647–1651.

109. Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2015). Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature *519*, 106–109.

110. Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J. Mol. Biol. *222*, 311–333.

111. Kjaergaard, M., and Poulsen, F.M. (2011). Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. J. Biomol. NMR *50*, 157–165.

112. Gil, S., Hošek, T., Solyom, Z., Kümmerle, R., Brutscher, B., Pierattelli, R., and Felli, I.C. (2013). NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. Angew. Chem. *125*, 12024–12028.

113. Bermel, W., Bertini, I., Felli, I., Piccioli, M., and Pierattelli, R. (2006). 13C-detected protonlessNMR spectroscopy of proteins in solution. Prog. Nucl. Magn. Reson. Spectrosc. *48*, 25–45.

114. Schneider, R., Maurin, D., Communie, G., Kragelj, J., Hansen, D.F., Ruigrok, R.W.H., Jensen, M.R., and Blackledge, M. (2015). Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. J. Am. Chem. Soc. *137*, 1220–1229.

115. Charlier, C., Bouvignies, G., Pelupessy, P., Walrant, A., Marquant, R., Kozlov, M., De Ioannes, P., Bolik-Coulon, N., Sagan, S., Cortes, P., et al. (2017). Structure and dynamics of an intrinsically disordered protein region that partially folds upon binding by chemical-exchange NMR. J. Am. Chem. Soc. *139*, 12219–12227.

116. Madhurima, K., Nandi, B., Munshi, S., Naganathan, A.N., and Sekhar, A. (2023). Functional regulation of an intrinsically disordered protein via a conformationally excited state. Sci. Adv. *9*, eadh4591.

117. Dedmon, M.M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C.M. (2005). Mapping long-range interactions in α-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J. Am. Chem. Soc. *127*, 476–477.

118. Kümmerer, F., Orioli, S., Harding-Larsen, D., Hoffmann, F., Gavrilov, Y., Teilum, K., and Lindorff-Larsen, K. (2021). Fitting side-chain NMR relaxation data using molecular simulations. J. Chem. Theor. Comput. *17*, 5262–5275.

119. Olsson, S., and Noé, F. (2017). Mechanistic models of chemical exchange induced relaxation in protein NMR. J. Am. Chem. Soc. *139*, 200–210.

120. Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. *12*, 54–60.

121. Fuxreiter, M. (2020). Fuzzy protein theory for disordered proteins. Biochem. Soc. Trans. *48*, 2557–2564.

122. Milles, S., Mercadante, D., Aramburu, I.V., Jensen, M.R., Banterle, N., Koehler, C., Tyagi, S., Clarke, J., Shammas, S.L., Blackledge, M., et al. (2015). Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. Cell *163*, 734–745.

123. Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. Nature *555*, 61–66.

124. Schneider, R., Blackledge, M., and Jensen, M.R. (2019). Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. Curr. Opin. Struct. Biol. *54*, 10–18.

125. Hansen, D.F., and Led, J.J. (2006). Determination of the geometric structure of the metal site in a blue copper protein by paramagnetic NMR. Proc. Natl. Acad. Sci. USA *103*, 1738–1743.

126. Fawzi, N.L., Ying, J., Ghirlando, R., Torchia, D.A., and Clore, G.M. (2011). Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. Nature *480*, 268–272.

127. Mittag, T., and Pappu, R.V. (2022). A conceptual framework for understanding phase separation and addressing open questions and challenges. Mol. Cell *82*, 2201–2214.

128. Murthy, A.C., and Fawzi, N.L. (2020). The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. J. Biol. Chem. *295*, 2375–2384.

129. Reichheld, S.E., Muiznieks, L.D., Keeley, F.W., and Sharpe, S. (2017). Direct observation of structure and dynamics during phase separation of an elastomeric protein. Proc. Natl. Acad. Sci. USA *114*, E4408–E4415.

130. Conicella, A.E., Zerze, G.H., Mittal, J., and Fawzi, N.L. (2016). ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. Structure *24*, 1537–1549.

131. Brady, J.P., Farber, P.J., Sekhar, A., Lin, Y.-H., Huang, R., Bah, A., Nott, T.J., Chan, H.S., Baldwin, A.J., Forman-Kay, J.D., and Kay, L.E. (2017). Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. Proc. Natl. Acad. Sci. USA *114*, E8194–E8203.

132. Kim, T.H., Tsang, B., Vernon, R.M., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2019). Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. Science *365*, 825–829.

133. Kim, T.H., Payliss, B.J., Nosella, M.L., Lee, I.T.W., Toyama, Y., Forman-Kay, J.D., and Kay, L.E. (2021). Interaction hot spots for phase separation revealed by NMR studies of a CAPRIN1 condensed phase. Proc. Natl. Acad. Sci. USA *118*, e2104897118.

134. Monahan, Z., Ryan, V.H., Janke, A.M., Burke, K.A., Rhoads, S.N., Zerze, G.H., O'Meally, R., Dignon, G.L., Conicella, A.E., Zheng, W., et al. (2017). Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. EMBO J. *36*, 2951–2967.

135. Yuwen, T., Brady, J.P., and Kay, L.E. (2018). Probing conformational exchange in weakly interacting, slowly exchanging protein systems via off-resonance $R_{1\rho}$ experiments: Application to studies of protein phase separation. J. Am. Chem. Soc. *140*, 2115–2126.

136. Robustelli, P., Ibanez-de-Opakua, A., Campbell-Bezat, C., Giordanetto, F., Becker, S., Zweckstetter, M., Pan, A.C., and Shaw, D.E. (2022). Molecular basis of small-molecule binding to α-synuclein. J. Am. Chem. Soc. *144*, 2501–2510.

137. Ban, D., Iconaru, L.I., Ramanathan, A., Zuo, J., and Kriwacki, R.W. (2017). A small molecule causes a population shift in the conformational landscape of an intrinsically disordered protein. J. Am. Chem. Soc. *139*, 13692–13700.

138. Heller, G.T., Shukla, V.K., Figueiredo, A.M., and Hansen, D.F. (2023). Small-molecule binding to an intrinsically disordered protein revealed by experimental NMR 19F transverse spin-relaxation. bioRxiv, 539297. https://doi.org/10.1001/2023.05.08.539824.

139. Zhang, O., Haghighatlari, M., Li, J., Liu, Z.H., Namini, A., Teixeira, J.M.C., Forman-Kay, J.D., and Head-Gordon, T. (2023). Learning to evolve structural ensembles of unfolded and disordered proteins using experimental solution data. J. Chem. Phys. *158*, 174113.

140. Li, J., Bennett, K.C., Liu, Y., Martin, M.V., and Head-Gordon, T. (2020). Accurate prediction of chemical shifts for aqueous protein structure on "real world" data. Chem. Sci. *11*, 3180–3191.

141. Carrara, E.A., Pagliari, F., and Nicolini, C. (1993). Neural networks for the peak-picking of nuclear magnetic resonance spectra. Neural Network. *6*, 1023–1032.

142. Corne, S.A., Johnson, A., and Fisher, J. (1992). An artificial neural network for classifying cross peaks in two-dimensional NMR spectra. J. Magn. Reson. *100*, 256–266.

143. Kjær, M., and Poulsen, F.M. (1991). Identification of 2D 1H NMR antiphase cross peaks using a neural network. J. Magn. Reson. *94*, 659–663.

144. Hare, B.J., and Prestegard, J.H. (1994). Application of neural networks to automated assignment of NMR spectra of proteins. J. Biomol. NMR *4*, 35–46.

145. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

146. Berjanskii, M.V., and Wishart, D.S. (2008). Application of the random coil index to studying protein flexibility. J. Biomol. NMR *40*, 31–48.

147. Karunanithy, G., Mackenzie, H.W., and Hansen, D.F. (2021). Virtual homonuclear decoupling in direct detection nuclear magnetic resonance experiments using deep neural networks. J. Am. Chem. Soc. *143*, 16935–16942.

148. Moutzouri, P., Cordova, M., Simões de Almeida, B., Torodii, D., and Emsley, L. (2023). Two-dimensional pure isotropic proton solid state NMR. Angew. Chem., Int. Ed. Engl. *62*, e202301963.

149. Hansen, D.F. (2019). Using deep neural networks to reconstruct non-uniformly sampled NMR spectra. J. Biomol. NMR *73*, 577–585.

150. Karunanithy, G., and Hansen, D.F. (2021). FID-Net: A versatile deep neural network architecture for NMR spectral reconstruction and virtual decoupling. J. Biomol. NMR *75*, 179–191.

151. Qu, X., Huang, Y., Lu, H., Qiu, T., Guo, D., Agback, T., Orekhov, V., and Chen, Z. (2020). Accelerated nuclear magnetic resonance spectroscopy with deep learning. Angew. Chem. *132*, 10383–10386.

152. Luo, J., Zeng, Q., Wu, K., and Lin, Y. (2020). Fast reconstruction of non-uniform sampling multidimensional NMR spectroscopy via a deep neural network. J. Magn. Reson. *317*, 106772.

153. Bostock, M., and Nietlispach, D. (2017). Compressed sensing: Reconstruction of non-uniformly sampled multidimensional NMR data. Concepts Magn. Reson. *46A*, e21438.

154. Miljenović, T.M., Jia, X., and Mobli, M. (2018). Nonuniform Sampling in Biomolecular NMR. In Modern Magnetic Resonance, G.A. Webb, ed. (Springer International Publishing), pp. 2035–2054.

155. Robson, S., Arthanari, H., Hyberts, S.G., and Wagner, G. (2019). Nonuniform sampling for NMR spectroscopy. In Methods in Enzymology (Elsevier), pp. 263–291.

156. Tugarinov, V., Kay, L.E., Ibraghimov, I., and Orekhov, V.Y. (2005). High-resolution four-dimensional 1H−13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J. Am. Chem. Soc. *127*, 2767–2775.

157. Żerko, S., Byrski, P., Włodarczyk-Pruszyński, P., Górka, M., Ledolter, K., Masliah, E., Konrat, R., and Koźmiński, W. (2016). Five and four dimensional experiments for robust backbone resonance assignment of large intrinsically disordered proteins: application to Tau3x protein. J. Biomol. NMR *65*, 193–203.

Structure
Review

CellPress
OPEN ACCESS

158. Jaravine, V., Ibraghimov, I., and Orekhov, V.Y. (2006). Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. Nat. Methods *3*, 605–607.

159. Hyberts, S.G., Milbradt, A.G., Wagner, A.B., Arthanari, H., and Wagner, G. (2012). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. J. Biomol. NMR *52*, 315–327.

160. Ying, J., Delaglio, F., Torchia, D.A., and Bax, A. (2017). Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. J. Biomol. NMR *68*, 101–118.

161. Hogben, H.J., Krzystyniak, M., Charnock, G.T.P., Hore, P.J., and Kuprov, I. (2011). Spinach – A software library for simulation of spin dynamics in large spin systems. J. Magn. Reson. *208*, 179–194.

162. Bak, M., Rasmussen, J.T., and Nielsen, N.C. (2011). SIMPSON: a general simulation program for solid-state NMR spectroscopy. J. Magn. Reson. *213*, 366–400.

163. Bloch, F., and Siegert, A. (1940). Magnetic resonance for nonrotating fields. Phys. Rev. *57*, 522–527.

164. Ying, J., Li, F., Lee, J.H., and Bax, A. (2014). 13Cα decoupling during direct observation of carbonyl resonances in solution NMR of isotopically enriched proteins. J. Biomol. NMR *60*, 15–21.

165. Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. *131*, 373–378.

166. Duma, L., Hediger, S., Lesage, A., and Emsley, L. (2003). Spin-state selection in solid-state NMR. J. Magn. Reson. *164*, 187–195.

167. Bermel, W., Bertini, I., Felli, I.C., Matzapetakis, M., Pierattelli, R., Theil, E.C., and Turano, P. (2007). A method for Cα direct-detection in protonless NMR. J. Magn. Reson. *188*, 301–310.

168. Bermel, W., Bertini, I., Duma, L., Felli, I.C., Emsley, L., Pierattelli, R., and Vasos, P.R. (2005). Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew. Chem., Int. Ed. Engl. *44*, 3089–3092.

169. Yang, Z., Zheng, X., Gao, X., Zeng, Q., Yang, C., Luo, J., Zhan, C., and Lin, Y. (2023). Deep learning methodology for obtaining ultraclean pure shift proton nuclear magnetic resonance spectra. J. Phys. Chem. Lett. *14*, 3397–3402.

170. Zangger, K. (2015). Pure shift NMR. Prog. Nucl. Magn. Reson. Spectrosc. *86*, 1–20.

171. Meyer, N.H., and Zangger, K. (2013). Simplifying proton NMR spectra by instant homonuclear broadband decoupling. Angew. Chem. *52*, 7143–7146.

172. Vallurupalli, P., Sekhar, A., Yuwen, T., and Kay, L.E. (2017). Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer. J. Biomol. NMR *67*, 243–271.

173. Tugarinov, V., Ceccon, A., and Clore, G.M. (2022). NMR methods for exploring 'dark'states in ligand binding and protein-protein interactions. Prog. Nucl. Magn. Reson. Spectrosc. *128*, 1–24.

174. Boisbouvier, J., and Kay, L.E. (2018). Advanced isotopic labeling for the NMR investigation of challenging proteins and nucleic acids. J. Biomol. NMR *71*, 115–117.

175. Klukowski, P., Augoff, M., Zięba, M., Drwal, M., Gonczarek, A., and Walczak, M.J. (2018). NMRNet: a deep learning approach to automated peak picking of protein NMR spectra. Bioinformatics *34*, 2590–2597.

176. Li, D.-W., Hansen, A.L., Yuan, C., Bruschweiler-Li, L., and Brüschweiler, R. (2021). DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra. Nat. Commun. *12*, 5229.

177. Karunanithy, G., Yuwen, T., Kay, L.E., and Hansen, D.F. (2022). Towards autonomous analysis of chemical exchange saturation transfer experiments using deep neural networks. J. Biomol. NMR *76*, 75–86.

178. NMRO. NMROnline – ELECTRO Cloud-Based Analysis Software. https://nmronline.uk.

179. baron de Fourier, J.B.J. (1822). Theorie analytique de la chaleur (Firmin Didot).

180. Wüthrich, K. (1990). Protein structure determination in solution by NMR spectroscopy. J. Biol. Chem. *265*, 22059–22062.

181. Karamanos, T.K., and Matthews, S. (2023). Biomolecular NMR in the AI-assisted structural biology era: Old tricks and new opportunities. Biochim. Biophys. Acta, Proteins Proteomics, 140949.