Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Chemische Biologie
der Technische Universität Dortmund

Characterization of protein structure and dynamics using solution- and solid-state NMR

Hanna Kavaleuskaya
aus
Minsk, Belarus

19 March, 2026

Kontakt

MSc. Hanna Kavaleuskaya
Fakultät für Chemie und Chemische Biologie
Technische Universität Dortmund
Otto-Hahn Str. 4a
44227 Dortmund
E-mail: hanna.kavaleuskaya@tu-dortmund.de
ORCID: 0000-0002-3636-2068

Abstract

Proteins execute biological function through structure and dynamics, often sampling multiple conformations and containing intrinsically disordered regions. Obtaining residue-specific information remains difficult for large, poorly soluble, heterogeneous, or highly dynamic systems, where many established methods lose applicability or resolution. This thesis develops strategies for studying such targets using solution- and magic-angle-spinning (MAS) solid-state NMR, complemented where appropriate by computational and evolutionary analyses.

The work is organized around three conceptually linked objectives. First, it addresses a major limitation of perdeuteration-based ^{1}H-detected MAS NMR, insufficient H/D back exchange that obscures solvent-shielded amides, by evaluating cell-extract-based selective deuteration to achieve high amide protonation in H_2O while maintaining favorable spectra. Second, it defines quantitative criteria for MAS sample preparation by benchmarking how solution, sedimented, and microcrystalline states affect resolution, sensitivity, and stability, and uses these comparisons to build a practical workflow for generating and validating backbone assignments from combined solution and MAS datasets. Third, it applies residue-resolved solution NMR to the intrinsically disordered N-terminal domain of cGAS to quantitatively assess conformational dynamics of apo NTD and NTD upon binding to various DNA molecules, and to place these findings in an evolutionary context.

Together, these studies show how labeling, sample state, and analysis choices jointly determine NMR information content in challenging biomolecular systems. Targeted labeling recovers otherwise inaccessible backbone sites without refolding. Sediments and microcrystals display distinct, measurable trade-offs in spectral quality and experimental robustness. For cGAS, the results support a modular organization of the N-terminal domain and indicate that DNA-induced responses are dominated by physical properties of DNA rather than its sequence. These advances broaden the practical toolbox for NMR studies of protein structure.

Zusammenfassung

Proteine erfüllen biologische Funktionen durch ihre Struktur und Dynamik, wobei sie häufig mehrere Konformationen einnehmen und oft intrinsisch ungeordnete Regionen enthalten. Die Gewinnung residuespezifischer Informationen bleibt für große, schlecht lösliche, heterogene oder stark dynamische Systeme schwierig, da viele etablierte Methoden dort an Anwendbarkeit oder Auflösung verlieren. Diese Dissertation entwickelt Strategien zur Untersuchung solcher Targets mittels Lösungs-NMR und Festkörper-NMR unter Magic-Angle-Spinning (MAS), ergänzt, wo sinnvoll, durch rechnergestützte und evolutionäre Analysen.

Die Arbeit ist um drei konzeptionell verknüpfte Ziele organisiert. Erstens adressiert sie eine zentrale Einschränkung der perdeuterierungsbasierten, ^{1}H-detektierten MAS-NMR—eine unzureichende H/D-Rückaustauschrate, die lösungsmittelabgeschirmte Amide verdeckt—indem eine selektive Deuterierung auf Basis von Zellextrakten evaluiert wird, um eine hohe Amid-Protonierung in H_2O bei gleichzeitig günstigen Spektraleigenschaften zu erreichen. Zweitens definiert sie quantitative Kriterien für die MAS-Probenpräparation, indem systematisch verglichen wird, wie sich Lösung, Sediment und mikrokristalline Zustände auf Auflösung, Empfindlichkeit und Stabilität auswirken, und nutzt diese Vergleiche, um einen praxisnahen Workflow zur Erzeugung und Validierung von Rückgrat-Zuordnungen aus kombinierten Lösungs- und MAS-Datensätzen zu entwickeln. Drittens wendet sie residuespezifische Lösungs-NMR auf die intrinsisch ungeordnete N-terminale Domäne von cGAS an, um die Konformationsdynamik der apo-NTD sowie der NTD bei Bindung an verschiedene DNA-Moleküle quantitativ zu erfassen und diese Befunde in einen evolutionären Kontext einzuordnen.

Zusammen zeigen diese Studien, wie Markierungsschema, Probenzustand und Auswerteentscheidungen gemeinsam den Informationsgehalt der NMR in anspruchsvollen biomolekularen Systemen bestimmen. Zielgerichtete Markierungsstrategien machen ansonsten unzugängliche Rückgratpositionen sichtbar, ohne Refolding-Schritte zu erfordern. Sedimente und Mikrokristalle zeigen unterschiedliche, messbare Kompromisse zwischen spektraler Qualität und experimenteller Robustheit. Für cGAS stützen die Ergebnisse eine modulare Organisation der N-terminalen Domäne und zeigen, dass DNA-induzierte Antworten vor allem von den physikalischen Eigenschaften der DNA und weniger von ihrer Sequenz geprägt sind. Diese Fortschritte erweitern das praktische Methodenspektrum für NMR-Studien der Proteinstruktur.

Contents

List of Acronyms

BEAST Bayesian Evolutionary Analysis Sampling Trees. 70

BMRB Biological Magnetic Resonance Data Bank. 69

CP Cross-Polarization. 21

CPMG Carr-Purcell-Meiboom-Gill. 20

CSA Chemical Shift Anisotropy. 10

CV Column Volume. 29

DNP Dynamic Nuclear Polarization. 21

DSS 4,4-dimethyl-4-silapentane-1-sulfonic acid. 13

EDTA Ethylenediaminetetraacetic Acid. 24

FID Free Induction Decay. 7

FT Fourier Transformation. 9

GA Glutaraldehyde. 46

HSQC Heteronuclear Single Quantum Coherence. 25

iFD Inverse Fractional Deuteration. 27

INEPT Insensitive Nuclei Enhanced by Polarization Transfer. 22

KDE Kernel Density Estimate. 47

MAS Magic Angle Spinning. 21

NMR Nuclear Magnetic Resonance. 3

NOE Nuclear Overhauser Effect. 16

NUS Non-Uniform Sampling. 27

OC Optimal Control. 22

POF Product Operator Formalism. 8

ppm Parts per million. 13

PRE Paramagnetic Relaxation Enhancement. 24

RD Relaxation Dispersion. 20

RDC Residual Dipolar Coupling. 15

RF Radio Frequency. 6

ROI Region of Interest. 30

SDS Sodium Dodecyl Sulphate. 30

SNR Signal-to-Noise Ratio. 21

ssNMR Solid-state NMR. 21

TMS Tetramethylsilane. 13

Publications

1. **Aucharova, H.**, and Linser, R. "Assignment of the N-terminal domain of mouse cGAS." *Biomolecular NMR Assignments* **19**(1), 35–39 (2025). doi: 10.1007/s12104-024-10213-2.

2. Bell, D., Lindemann, F., Gerland, L., **Aucharova, H.**, Klein, A., Friedrich, D., Hiller, M., Grohe, K., Meier, T., van Rossum, B., Diehl, A., Hughes, J., Mueller, L.J., Linser, R., Miller, A.F., and Oschkinat, H. "Sedimentation of large, soluble proteins up to 140 kDa for ^{1}H-detected MAS NMR and ^{13}C DNP NMR – practical aspects." *Journal of Biomolecular NMR* **78**(3), 179–192 (2024). doi: 10.1007/s10858-024-00444-9.

3. **Aucharova, H.**, Klein, A., Gomez, S.M., Söldner, B., Vasa, S.K., and Linser, R. "Protein deuteration via algal amino acids to circumvent proton back-exchange for ^{1}H-detected solid-state NMR." *Chemical Communications* **60**(22), 3083–3086 (2024). doi: 10.1039/d4cc00213j.

4. **Aucharova, H.**, Blokhin, D., and Linser, R. "DNA-binding properties of a disordered domain of cGAS." *Manuscript in preparation* (2026).

Chapter 1

Introduction

This chapter introduces the basic physical, mathematical, and chemical concepts underlying the methods used in the work. The first part of the chapter focuses on theoretical basics followed by practical aspect of biomolecular Nuclear Magnetic Resonance (NMR). Elements of protein dynamics as well as brief history and key aspects of solid-state MAS NMR applied to biomolecules are covered in second part of the chapter.

1.1 Place of structural biology

Proteins drive cellular processes through their structure and dynamics. The three-dimensional architecture of proteins emerges from their primary amino acid sequences, resulting in diverse protein types. Historically, proteins were considered to exist as single, stable, and rigid entities. However, in solution or when tethered to membranes, proteins exhibit dynamic behavior, transitioning between multiple conformations ranging from highly ordered to completely disordered states. The multidimensional energy landscape determines the thermodynamics and kinetics of these conformational transitions. The discovery of fold-switching proteins, prion proteins, and the phenomenon of folding upon binding in certain IDPs/IDRs has shown that protein dynamics are integral to nearly all protein functions and are evident at multiple levels, including conformational dynamics, conformational changes, protein folding, and intermolecular interactions. A central challenge of structural biology nowadays is the dynamic characterization of conformational ensembles.

In recent decades, X-ray crystallography and NMR spectroscopy have been fundamental techniques in structural biology, enabling the determination of atomic- or near-atomic-resolution three-dimensional structures and protein structural ensembles. Technological advancements, including improved crystallization protocols, increased magnetic field strengths and rotor spinning speeds, and the development of specialized NMR experiments have broadened the spectrum of molecular targets, facilitating the structural analysis of complex proteins and multi-domain assemblies that were previously inaccessible.

To correlate protein architecture with biological activity, structures and/or structural ensembles derived with X-ray, cryo-EM, NMR methods must be enriched by functional assays and computational simulations that expose the conformational transitions proteins undergo during catalysis, allosteric regulation, or molecular recognition. Additionally, evolutionary data provide valuable insights into how protein variability influences stability, alters active site geometry, or redirects protein–protein interactions. Comprehensive analyses of proteins evolution, structure, dynamics, and biological functions are essential for advanced drug design, elucidation of disease mechanisms, and protein engineering for biomedical and industrial applications.

This thesis employs NMR spectroscopy to investigate protein structures and their associated dynamics, and incorporates molecular evolution to contextualize structural information within an evolutionary framework.

1.2 Fundamental Techniques in NMR Spectroscopy

Throughout this chapter we use the following notation. Italic roman symbols (e.g. l, m_l, I) denote scalar quantities such as quantum numbers and eigenvalues. Bold upright symbols (e.g. $\mathbf{L}$, $\mathbf{I}$) denote vector quantities such as orbital and nuclear spin angular momenta, and their magnitudes are written as scalars with absolute value bars, e.g. $|\mathbf{L}|$. Operators are written as bold italic symbols with a hat, for example $\hat{\boldsymbol{L}}$ for the orbital angular momentum operator and $\hat{\boldsymbol{I}}$, $\hat{\boldsymbol{S}}$ for spin operators. A summary of the quantum-mechanical parameters describing the spin system is given in Supplementary Table A.1.

1.2.1 Quantum mechanics of spin systems

In quantum mechanics, the state of a nucleus is defined by a set of quantum numbers, which describe its intrinsic and angular momentum properties. These quantum numbers determine how nucleons (protons and neutrons) behave within the nucleus, their energy levels, and their interactions with external magnetic fields, which are crucial for NMR spectroscopy.

Energy states and population distribution are the fundamental subjects of any spectroscopic technique. The energy difference between energy states gives rise to the frequency of the spectra, whereas intensities of the spectral peaks are proportional to the population difference of the states.

NMR spectroscopy exploits behavior of nuclei with a spin angular momentum in the presence of an external magnetic field. The key requirement for a nucleus to be NMR-active is having a non-zero nuclear spin quantum number I. Nuclei with an even number of protons Z and neutrons N have $I = 0$, for example, ^{12}C. These kinds of nuclei will not give rise to an NMR signal. For nuclei with $I \neq 0$, energy states arise due to the interaction of nuclear spin angular momentum $\mathbf{I}$ with the applied magnetic field $\mathbf{B}_0$. The total nuclear spin angular momentum has a fixed magnitude given by:

$$|\mathbf{I}| = \hbar\sqrt{I(I+1)} \tag{1.1}$$

, where I is the nuclear spin quantum number, and $\hbar$ is the reduced Planck constant ($\hbar = h/2\pi$). The value of I depends on the mass and charge distribution of the nucleus and can take values $I = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$. Isotopes such as ^{1}H, ^{13}C, and ^{15}N, which are particularly important in biological applications, have a spin-$\frac{1}{2}$ nucleus. While ^{1}H is abundant in nature, the ^{13}C and ^{15}N isotopes need to be introduced artificially, as their natural abundance is typically insufficient for detection by NMR spectroscopy.

While the total spin angular momentum $|\mathbf{I}|$ is a fixed quantity, its measurable component in a given direction is determined by the magnetic quantum number m_I. The Z-component of the spin angular momentum is given by:

$$I_z = \hbar m_I \tag{1.2}$$

, where m_I can take values of I, $I-1$, ..., $-I+1$, $-I$, resulting in a total of $2I+1$ discrete states. This equation establishes that the spin angular momentum is quantized in space, meaning the nuclear spin can only have specific orientations relative to an external field. However, in the absence of magnetic field, all possible orientations of nuclear spin states have equal energy (i.e., spin state degeneracy).

1.2.2 Nuclear spin populations

When a nucleus is placed in a magnetic field $\mathbf{B}_0$, the Zeeman effect causes the degeneracy of spin states to be lifted, creating discrete energy levels based on m_I. For a spin-$\frac{1}{2}$ nucleus, the two allowed values of m_I are:

$m_I = +\frac{1}{2}$ ("up") $\rightarrow$ Parallel to $\mathbf{B}_0$

$m_I = -\frac{1}{2}$ ("down") $\rightarrow$ Antiparallel to $\mathbf{B}_0$

These orientations are also called α- and β-states. These states are not static; rather, the nuclear spin continuously precesses around the magnetic field direction due to its magnetic moment μ. (Fig.1.1)

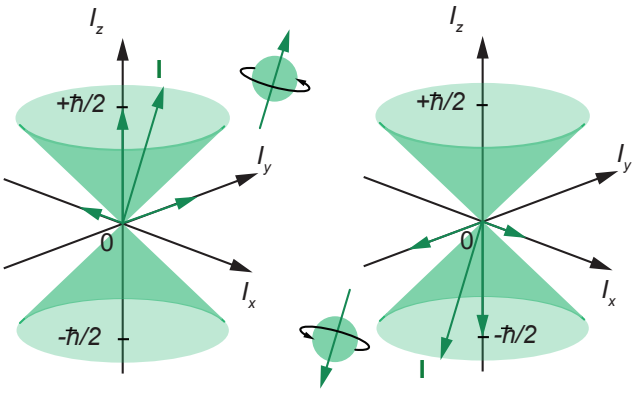


Figure 1.1: Orientation of nuclear spin angular momentum **I** with spin-$\frac{1}{2}$ and its Z-component, I_z. The vectors represent the angular momentum rotating about the external magnetic field, whose direction is along the Z-axis of the laboratory frame.

1.2.3 Larmor frequency

The potential energy E of a magnetic moment μ in an external magnetic field $\mathbf{B_0}$ is given by:

$$E = \boldsymbol{\mu} \cdot \mathbf{B_0} = -\gamma \hbar m_I B_0 \tag{1.3}$$

For each spin state, there is energy associated with it:

$$E_\alpha = -\tfrac{1}{2}\mu_z B_0 = -\tfrac{1}{2}\hbar\gamma B_0 \qquad E_\beta = +\tfrac{1}{2}\mu_z B_0 = +\tfrac{1}{2}\hbar\gamma B_0 \tag{1.4}$$

This energy difference between states determines the resonance (Larmor) frequency in NMR:

$$\Delta E_{\alpha\to\beta} = E_\beta - E_\alpha = \tfrac{1}{2}\hbar\gamma B_0 - \left(-\tfrac{1}{2}\hbar\gamma B_0\right) = \hbar\gamma B_0 = \left(\tfrac{h}{2\pi}\right)\gamma B_0 \tag{1.5}$$

Using Planck's relation, $\Delta E = h\nu$, Equation (1.6) can be rearranged to:

$$\nu_{\alpha\to\beta} = \frac{\Delta E_{\alpha\to\beta}}{h} = \frac{\gamma B_0}{2\pi} \tag{1.6}$$

In an applied magnetic field B_0, the Larmor frequency $\nu_{\alpha\to\beta}$ is determined by the gyromagnetic ratio γ. In the NMR literature, the frequency ν is often replaced by the angular frequency $\omega = 2\pi\nu$, which is expressed in $\mathrm{rad\,s^{-1}}$ units. Hence, the expression for the Larmor frequency becomes as follows:

$$\omega_0 = -\gamma B_0 \tag{1.7}$$

The Larmor frequency determines the resonance condition for the nuclei, as Radio Frequency (RF) pulses must match this frequency to excite nuclear spins. Accurate knowledge of the Larmor frequency for the nuclei under investigation is essential for defining the rotating frame, which is required for all NMR physics calculations. Chemical shifts are referenced to the Larmor frequency because they are defined by the difference between the observed resonance and the Larmor frequency.

1.2.4 Boltzmann distribution

The statistical behavior of nuclear spins and their relative populations in applied magnetic field at thermal equilibrium are described by Boltzmann distribution:

$$\frac{N_\alpha}{N_\beta} = e^{-\left(\frac{\Delta E}{kT}\right)} \tag{1.8}$$

, where ΔE is the energy difference between states, T is the absolute temperature in Kelvins, $k = 1.38{\times}10^{-23}\,\mathrm{J\,K^{-1}}$ is the Boltzmann constant. This equation indicates that both the energy difference between the transition states and the population difference of the states increase with the magnetic field strength. Furthermore, the population difference exhibits a temperature dependence. At room temperature, the population of the β-state is slightly lower than that of the α-state, resulting in a very small net bulk magnetization $\mathbf{M}_0$. Since detectable NMR signal arises from precession of bulk magnetization vector $\mathbf{M}_0$, the relatively small magnitude of $\mathbf{M}_0$ makes NMR a very insensitive spectroscopic method.

1.2.5 Rotating frame

Key information about chemical structure and molecular dynamics, such as chemical shifts, J-couplings, and other spin interactions, occurs on the kHz or sub-kHz scale (Supplementary Table A.3). To analyze differences within the desired frequency range, the common contribution to the Larmor frequency (the Zeeman interaction) can be effectively removed from the theoretical treatment. This is achieved by introducing a new coordinate system called the rotating frame. In the rotating frame, the XY-plane rotates at or near the Larmor frequency ω_0 about the laboratory Z-axis. This transformation effectively eliminates rapid precession, enabling observation of the individual spin system's evolution relative to the common basis.

1.2.6 Effects of the $\mathbf{B}_1$ field

When a macroscopic sample is placed in a static magnetic field $\mathbf{B}_0$, the nuclear spins partially align with the field, generating a small but measurable bulk magnetization vector $\mathbf{M}_0$. This magnetization has three Cartesian components, M_x, M_y, and M_z, corresponding to the X-, Y-, and Z-axes of the laboratory frame. At thermal equilibrium, only the longitudinal component M_z is non-zero. Applying an RF field $\mathbf{B}_1$ in the XY-plane tips the bulk magnetization away from the Z-axis. The $\mathbf{B}_1$ field, generated by an oscillating current in the probe coil, induces transitions between spin states and produces transverse magnetization. The resulting transverse components M_{xy} precess about the Z-axis at the Larmor frequency, inducing a voltage in the detection coil. This precessing transverse magnetization forms the observable NMR signal, known as the Free Induction Decay (FID). The frequency of rotation induced by the pulse is determined by:

$$\omega_1 = -\gamma \mathbf{B}_1 \tag{1.9}$$

By modulating the amplitude and duration of $\mathbf{B}_1$, the bulk magnetization can be rotated anywhere in the plane perpendicular to the axis of the applied $\mathbf{B}_1$ field in the rotating frame. If $\mathbf{B}_1$ is turned on and then off as $\mathbf{M}_0$ moves from the Z-axis

to the XY-plane, this is called a 90° pulse. The duration of this pulse is called the 90° pulse length, and the $\mathbf{B}_1$ field amplitude is expressed as the effective field, the frequency with which $\mathbf{B}_1$ field rotates a given magnetization. The pulse power for a hard (short) 90° pulse is typically about 50 W for protons and > 100 W for heteronuclei (^{13}C, ^{15}N). Because heteronuclei have lower gyromagnetic ratios than protons, they require longer 90° pulse lengths at a given power.

1.2.7 Product operator formalism

Modern multidimensional NMR spectroscopy requires monitoring not only magnetization but also its phase, coherence order, and magnetization transfer between coupled spins. The Product Operator Formalism (POF) provides a compact algebraic framework in which spin dynamics are expressed in terms of spin operators and their products. Specifically, in POF, nuclear spin dynamics are described using the Cartesian spin operators $\hat{I}_x$, $\hat{I}_y$, and $\hat{I}_z$ the components of the spin angular momentum along the X-, Y-, and Z-axes. In the rotating frame, transverse magnetization detected in a conventional 1D experiment is represented by single-spin operators such as $\hat{I}_x$ or $\hat{I}_y$. These terms are called in-phase magnetization because all spins of a given type precess with the same phase and give rise to absorption-mode lines. Consequently, in-phase magnetization encodes primarily chemical shift information. In multi-spin systems, scalar J-coupling introduces correlations between spins. These correlations are described by two-spin product operators such as $2\hat{I}_z\hat{S}_z$, $2\hat{I}_x\hat{S}_z$, and $2\hat{I}_y\hat{S}_z$ for a coupled IS-spin system. Operators of the form $2\hat{I}_x\hat{S}_z$ and $2\hat{I}_y\hat{S}_z$ represent anti-phase magnetization: the transverse magnetization of spin I depends on the spin state of S, leading to doublets whose two components have opposite sign. Anti-phase terms are central in J-resolved and correlation spectra because they carry information on scalar couplings and enable magnetization transfer.

The behavior of these operators under rotations defines the concept of coherence order p. Terms with $p = 0$ are longitudinal contributions, such as $\hat{I}_z$ and $2\hat{I}_z\hat{S}_z$, represent populations or z-magnetization and do not give a directly observable NMR signal. Terms with $p = \pm 1$ are single-quantum coherences, for example $\hat{I}_x$, $\hat{I}_y$, $2\hat{I}_x\hat{S}_z$, and $2\hat{I}_y\hat{S}_z$. These correspond to transverse magnetization that precesses at the Larmor frequency and produces the FID. Higher coherence orders ($p = \pm 2, \pm 3, \ldots$) describe multiple-quantum coherences, which are not observed directly but can be created and converted back into single-quantum coherence by suitable pulse sequences. In this way, from the coherence order one can derive which parts of the magnetization can be stored along Z, which parts are observable, and how magnetization is transferred between spins during an NMR experiment.

To describe how operators change in time, it is convenient to introduce the Hamiltonian. In quantum mechanics, the Hamiltonian $\hat{\mathcal{H}}$ is the energy operator of the spin system. In the presence of a strong static magnetic field $\mathbf{B}_0$ along the Z-axis, the Hamiltonian for a single spin-$\frac{1}{2}$ nucleus is given by:

$$\hat{\mathcal{H}}_0 = -\gamma \mathbf{B}_0 \hat{I}_z \tag{1.10}$$

, where γ is the gyromagnetic ratio and $\hat{I}_z$ is the Z-component of the spin angular momentum operator $\mathbf{I}$. This Hamiltonian describes the Zeeman interaction, where the magnetization precesses around the Z-axis at the Larmor frequency. In the rotating frame, the effect of $\hat{\mathcal{H}}_0$ is removed, simplifying further analysis.

An RF field applied along the X- or Y-axis induces transitions between spin states. The Hamiltonian for an RF pulse along the X-axis follows:

$$\hat{\mathcal{H}}_{RF} = \omega_1 \hat{I}_x \tag{1.11}$$

, where ω_1 is the strength of the RF field. Within the product operator formalism, RF pulses, chemical-shift evolution and J-couplings are therefore all treated as well-defined transformations between operators with specific coherence orders. This makes POF an essential tool for analyzing and designing modern multidimensional NMR pulse sequences.

1.2.8 Relaxation theory

Bloch equations

An NMR signal has unique properties that cannot be understood purely by considering the classical torque equation for a magnetic moment in a magnetic field: finite linewidths and characteristic relaxation behavior. To account for these effects, Felix Bloch modified the classical precession equations for the magnetic moment μ to describe spin precession, relaxation, and interactions with external magnetic fields [1][2]. These equations incorporate two key relaxation times: (a) longitudinal (spin-lattice) relaxation (T_1), which dictates the recovery of magnetization along the Z-axis (M_z) to thermal equilibrium, affecting signal intensity in inversion recovery experiments; and (b) transverse (spin-spin) relaxation (T_2), which controls the decay of magnetization in the XY-plane (M_{xy}) due to spin-spin interactions, influencing spectral linewidth and resolution.

$$\begin{cases} \dfrac{d}{dt}\mu_X(t) = -\omega\,\mu_Y(t) - \dfrac{1}{T_2}\mu_X(t), \\ \dfrac{d}{dt}\mu_Y(t) = +\omega\,\mu_X(t) - \dfrac{1}{T_2}\mu_Y(t), \\ \dfrac{d}{dt}\mu_Z(t) = -\dfrac{1}{T_1}\big(\mu_Z(t) - \mu_{\text{eq}}\big) \end{cases} \quad \text{or} \quad \begin{cases} \dfrac{d}{dt}M_X(t) = -\omega\,M_Y(t) - \dfrac{M_X(t)}{T_2}, \\ \dfrac{d}{dt}M_Y(t) = +\omega\,M_X(t) - \dfrac{M_Y(t)}{T_2}, \\ \dfrac{d}{dt}M_Z(t) = -\dfrac{1}{T_1}\big(M_Z(t) - M_0\big) \end{cases} \tag{1.12}$$

These equations describe the Larmor precession of nuclear spins around the static field at Larmor frequency, with the Z-component of the bulk magnetization M_z independent of time, and M_z and M_y oscillating and decaying due to relaxation effects. When an RF field $\mathbf{B}_1$ is applied in the XY-plane, the effective field changes, and the magnetization precesses about the new effective field:

$$\mathbf{B}_{\text{eff}} = \mathbf{B}_0 + \frac{\omega}{\gamma} \tag{1.13}$$

Before 1966, most chemical applications of NMR were performed using continuous-wave spectrometers, which relied on the detection of the steady-state transverse magnetization as a function of the static magnetic field. The Bloch equations lost their central place in NMR theory when the continuous-wave method was displaced by the Fourier Transformation (FT) technique. Moreover, the equations are phenomenological, e.g. just describe relaxation instead of explaining or deriving it.

Redfield theory

Redfield theory is a semi-classical theory that postulates that relaxation in the sample originates from random fluctuations of a local magnetic field, and the strength of relaxation is determined by how these fluctuations correlate in time. Redfield theory derives T_1 and T_2 expressions by (1) splitting the Hamiltonian of a system into a static Zeeman Hamiltonian $\hat{\mathcal{H}}_0$ and time-dependent perturbation Hamiltonian $\hat{\mathcal{H}}_1(t)$, (2) introducing the density matrix $\hat{\sigma}$ with a split Hamiltonian, (3) assessment of density matrix time evolution via the Liouville-von Neumann (LvN) equation, (4) transformation into the interaction frame to remove Larmor precession, (5) expansion of the dynamic equation to second order using perturbation theory, (6) removing the first order term that represents fluctuations that average out for the whole sample, (7) rewriting the density matrix at thermal equilibrium, (8) application of a weak collision limit that assumes that the density matrix changes slowly compared to fluctuations, (9) transformation back to the laboratory frame, (10) rewriting the equation with the Redfield relaxation superoperator $\hat{\hat{\Gamma}}$, (11) rewriting the relaxation superoperator in spherical tensor form, (12) expansion of the fluctuating interaction Hamiltonian $\hat{\mathcal{H}}_1(t)$ to produce a correlation function C, and (13) FT of C, which defines the spectral density $J_(\omega)$, which quantifies the strength of molecular motion at frequency ω that drives transitions between Zeeman energy levels:

$$J(\omega) = \frac{2\tau_c}{1 + \omega_0^2\tau_c^2}. \tag{1.14}$$

In this form, the longitudinal and transverse relaxation rates depend only on the spectral density evaluated at 0, ω_0, and $2\omega_0$ as well as the sums or differences of Larmor frequencies for spin-spin interactions:

$$R_1 = \frac{1}{T_1} \propto J(\omega_0) + J(2\omega_0), \qquad R_2 = \frac{1}{T_2} \propto \tfrac{1}{2}J(0) + \tfrac{1}{4}J(\omega_0) + \tfrac{1}{4}J(2\omega_0), \tag{1.15}$$

with proportionality constants determined by the specific interaction (dipolar, Chemical Shift Anisotropy (CSA), or quadrupolar). These expressions show that T_1 and T_2 experiments are sensitive to slightly different timescales: R_1 relaxation rates depend only on $J(\omega_0)$ and $J(2\omega_0)$, the probability of molecular motions that fluctuate near the Larmor frequency, while the R_2 relaxation rate equation contains both probabilities for fast oscillations $J(\omega_0)$ and $J(2\omega_0)$ and a slow-oscillating term proportional to $J(0)$. As a result, R_2 rate is sensitive not only to fast ps-ns timescale motions that also influence R_1, but also slow motions that broaden the transverse magnetization through dephasing. Redfield relaxation theory is implemented in biomolecular NMR by measuring site-specific relaxation parameters and mapping them to spectral density functions that encode motion on specific timescales.

1.2.9 Secular approximation

The total NMR Hamiltonian contains several interactions, including Zeeman, dipolar, J-coupling, and CSA (see 1.3.2 Interactions in NMR spectroscopy). As mentioned earlier, in a strong magnetic field, the Larmor frequency of nuclear spins is much higher than most other interactions. Fast oscillations at frequencies much

higher than the interaction strength contribute little to the long-term evolution of the system and can therefore be ignored. To simplify the Hamiltonian of a system evolving under multiple forces and interactions, the secular approximation is used. Mathematically, the total Hamiltonian of a nuclear spin system can be represented as follows:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_Z + \hat{\mathcal{H}}_{int} \tag{1.16}$$

, where $\hat{\mathcal{H}}_Z = -\gamma \mathbf{B}_0 \hat{I}_z$ is the Zeeman Hamiltonian, and $\hat{H}_{\text{int}}$ represents various interactions. The secular approximation states that we may neglect terms for which

$$\omega_i \gg |\hat{\mathcal{H}}_{int}| \tag{1.17}$$

, because these terms oscillate too rapidly to contribute meaningfully to the average, long-term dynamics of the spin system. The secular approximation reduces the complexity of the Hamiltonian by retaining only terms that evolve slowly in the rotating frame. This method yields a simplified yet accurate representation of dominant spin interactions under high magnetic field conditions and forms the foundation for most NMR pulse sequence analyses.

1.3 Practical aspects of NMR

1.3.1 Basic elements of NMR pulse programs

All NMR experiments consist of sequences of basic elements of NMR pulse programs, which dictate signal generation and coherence evolution, and, eventually, define the experimental outcome.

(a) **RF-pulses** - the most common pulses used in NMR experiments. A 90° pulse rotates the magnetization vector from the Z-axis to the transverse XY-plane, while a 180° pulse inverts the magnetization along the Z-axis or enables refocussing it in the transverse plane. Selective pulses are designed to affect only a specific range of frequencies, allowing for selective excitation or inversion of certain spins.

(b) **Delays** - can be grouped into evolution periods and mixing times. Evolution periods are time intervals during which the magnetization evolves under the influence of chemical shifts, J-couplings, and relaxation processes. Evolution periods are crucial for encoding frequency information in multidimensional NMR. Mixing times are specific delays used to transfer magnetization between different spins, often through mechanisms like cross-relaxation (NOE) or coherence transfer (COSY, INEPT).

(c) **Gradient pulses** - used to dephase or rephase magnetization in a spatially dependent manner. Gradients are essential for coherence selection, artifact suppression, and solvent suppression.

(d) **Phase cycling** - involves systematically varying the phase of RF pulses and the receiver to select desired coherence pathways and suppress unwanted signals. Phase cycling is a key tool for artifact reduction and signal selection.

(e) **Decoupling** - involves continuous or pulsed irradiation of a specific nuclear spin to remove its coupling effects on other spins, simplifying the spectrum and enhancing resolution.

Groups of pulse program elements comprise periods. Multidimensional NMR experiments are usually comprised of five distinct periods: preparation, evolution, mixing, detection, and recovery. During the preparation, the magnetization is set in a specific state, typically using a 90° pulse to create transverse magnetization from the equilibrium magnetization. In the evolution period, magnetization evolves under the influence of chemical shifts and J-couplings in a series of increments to encode the indirect dimension (t_1). During the mixing period, the magnetization is transferred between different spins, which can occur either through scalar coupling (e.g., through bonds) or through dipolar coupling (e.g. through space). In the detection period, NMR signal in form of an FID is recorded as a function of time (t_2), the signal is then Fourier-transformed to obtain the frequency domain spectrum. Finally, the recovery period is required to restore the equilibrium magnetization. Its duration is determined by the T_1 relaxation time of the nuclei of interest.

In multi-dimensional experiments, the sequence is repeated multiple times, incrementing the time variables during each evolution period. As a result, a number

of FIDs is detected with progressively increasing time variables in the evolution periods. The acquired FIDs are then subjected to FT, producing a series of frequency domain spectra. The spectral intensities oscillate as functions of time, e.g. t_1, t_2. This oscillation constitutes an FID, often referred to as the indirect FID. Applying an FT of the indirect FIDs results in a frequency plot, showing an intensity for each pair of frequency variables. The third dimension represents the intensities in the spectrum.

1.3.2 Interactions in NMR spectroscopy

Chemical shift

The chemical shift is one of the fundamental parameters in NMR spectroscopy, describing the resonance frequency of a nucleus relative to a reference compound (e.g., Tetramethylsilane (TMS) or 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) for ^{1}H and ^{13}C NMR). It reflects the local electronic environment of an atomic nucleus. The chemical shift $\boldsymbol{\delta}$ is measured in Parts per million (ppm) and is defined as:

$$\boldsymbol{\delta} = \frac{\nu - \nu_{\text{ref}}}{\nu_{\text{ref}}} \times 10^6 \tag{1.18}$$

, where ν is the absolute resonance frequency of the sample, ν_{ref} is the absolute resonance frequency of a standard reference compound. The chemical shift arises from the local magnetic environment of the nucleus, which is affected by the surrounding electron cloud. When placed in an external magnetic field $\mathbf{B}_0$, the circulating electrons create an induced local magnetic field that opposes $\mathbf{B}_0$, altering the effective field experienced by the nucleus. The net effect of electronic shielding can be described using a shielding tensor σ by:

$$\nu = \frac{\gamma}{2\pi} B_0 (1 - \sigma) \tag{1.19}$$

The shielding is dependent on the distribution of electron density surrounding the nucleus. Since the distribution is spherically asymmetrical, $\boldsymbol{\sigma}$ usually has an anisotropic value with an orientational dependence. Mathematically, shielding tensor is a second-rank tensor written as a 3x3 matrix:

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \tag{1.20}$$

After diagonalization, diagonal elements of $\boldsymbol{\sigma}$ represent the principal shielding values along three mutually perpendicular axes in the principal axis system. The off-diagonal elements of the chemical shielding tensor are hence eliminated by transforming to the principal axis system where the tensor is diagonal, and by applying the secular approximation which removes fast-oscillating terms.

Several factors can contribute to the shielding constant: diamagnetic shielding $\boldsymbol{\sigma}_{\text{dia}}$, paramagnetic shielding $\boldsymbol{\sigma}_{\text{para}}$, and ring current effect. Diamagnetic shielding has an opposite direction to the $\mathbf{B}_0$ field and originates from a spherical electronic distribution. Paramagnetic shielding originates from a non-spherical electronic distribution, in which the induced local field has the same direction as $\mathbf{B}_0$. $\boldsymbol{\sigma}_{\text{dia}}$ and

$\boldsymbol{\sigma}_{\text{para}}$ have opposite contributions to the shielding constant, therefore some of the effects are canceled out.

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{dia}} + \boldsymbol{\sigma}_{\text{para}} \tag{1.21}$$

In protons, the large energy gap results in minimal paramagnetic shielding, even when bonding distorts the spherical electron distribution. This leads to a narrow chemical shift range, typically around 10 ppm. In contrast, paramagnetic contributions generally dominate over the diamagnetic term in heteronuclei, which accounts for the significantly broader chemical shift range observed for heteronuclei compared to ^{1}H. The ring current effect originates from delocalized electrons in $p-$orbitals moving between bonded atoms within an aromatic ring. When an aromatic ring is placed in a magnetic field $\mathbf{B}_0$ perpendicular to the ring plane, circulation of the π-electrons induces a ring current that generates a secondary magnetic field. This induced field opposes $\mathbf{B}_0$ above and below the ring center, but adds to $\mathbf{B}_0$ around the ring periphery in the ring plane. Consequently, aromatic protons at the ring edge experience a larger effective field and are shifted downfield. The ring current effect has a reduced influence on the ^{13}C chemical shifts of aromatic compounds because carbon nuclei are positioned near the region where the induced magnetic field changes direction between shielding and deshielding, resulting in a field that is nearly zero.

J-coupling

Scalar *J*-coupling is a fundamental interaction in NMR spectroscopy that originates from the indirect through-bond communication between two nuclear spins, I and S, mediated by the polarization of bonding electrons. When spin S aligns with the external magnetic field $\mathbf{B}_0$, it induces a polarization of the surrounding electron cloud, primarily within the *s*-orbitals. This polarization alters the electron density around spin I which is directly bonded to S, thereby facilitating spin interaction. As this effect requires the nuclei to be in Fermi contact (a magnetic interaction between an electron and an atomic nucleus), *J*-coupling is exclusively a through-bond interaction. The *J*-coupling Hamiltonian for two coupled spins I and S is:

$$\hat{\mathcal{H}}_J = 2\pi J\,(\mathbf{I} \cdot \mathbf{S}) \tag{1.22}$$

Expanding the dot product of $\mathbf{I} \cdot \mathbf{S}$:

$$\hat{\mathbf{H}}_J = 2\pi J \left(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y + \hat{I}_z\hat{S}_z\right) \tag{1.23}$$

, where J is the scalar coupling constant (in Hz). Some spin operator interaction terms can be removed to simplify calculations using secular approximation, provided the Zeeman interaction ($\omega_I - \omega_S$) is much larger than the *J*-coupling. This means, that terms oscillating at Larmor frequency ($\hat{I}_x\hat{S}_x$,$\hat{I}_y\hat{S}_y$) average out over time, and only the $\hat{I}_z\hat{S}_z$ term remains:

$$\hat{\mathcal{H}}_J = 2\pi J\,\hat{I}_z\hat{S}_z \tag{1.24}$$

The remaining $\hat{I}_z\hat{S}_z$ term causes energy level splitting, which leads to multiplets in solution-state NMR.

Although J-coupling is inherently anisotropic due to the asymmetric electronic environment, rapid molecular tumbling in solution averages it to an isotropic value. The J-coupling magnitude decreases significantly with the number of intervening bonds, with two-bond ($2J$) and three-bond ($3J$) couplings typically being at least an order of magnitude weaker than one-bond ($1J$) couplings.

Dipolar coupling

Dipolar coupling is a fundamental interaction in NMR that originates from the direct through-space interaction between nuclear spins (I and S) separated by an internuclear vector $\mathbf{r}$ (with $r = |\mathbf{r}|$). Each nuclear spin generates a local magnetic field, which influences the energy levels of nearby nuclei. This effect is described by the dipolar Hamiltonian:

$$\begin{aligned}\hat{\mathcal{H}}_D &= -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi r^3}\left[3(\hat{\mathbf{I}}\cdot\hat{\mathbf{r}})(\hat{\mathbf{S}}\cdot\hat{\mathbf{r}}) - (\hat{\mathbf{I}}\cdot\hat{\mathbf{S}})\right] \\ &= -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi r^3}\left(\frac{3\cos^2\theta - 1}{2}\right)\left[2\hat{I}_z\hat{S}_z - \left(\hat{I}_x\hat{S}_x + \hat{I}_y\hat{S}_y\right)\right]\end{aligned} \tag{1.25}$$

, where μ_0 is the magnetic permeability of free space, $\hbar$ is the reduced Planck constant, γ_I and γ_S are the gyromagnetic ratios of the interacting nuclei, $r = |\mathbf{r}|$ is the internuclear distance, and $\hat{\mathbf{r}} = \mathbf{r}/r$ is the unit vector along the internuclear axis. $\hat{\mathbf{I}}$ and $\hat{\mathbf{S}}$ are the spin angular momentum operators for the two nuclei, $\hat{I}_z\hat{S}_z$ – longitudinal term, $\hat{I}_x\hat{S}_x$, $\hat{I}_y\hat{S}_y$ – transverse terms of dipolar Hamiltonian. Transverse terms oscillate at high frequencies and are removed under the secular approximation. If we now express the internuclear vector $\mathbf{r}$ in spherical coordinates, project it onto the external field direction, and evaluate only frequency-independent secular terms of dipolar Hamiltonian, Equation (1.25) transforms into:

$$\hat{\mathcal{H}}_D^{\text{secular}} = -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi\, r^3}\left(\frac{3\cos^2\theta - 1}{2}\right)\hat{I}_z\hat{S}_z \tag{1.26}$$

The resonance frequency shift is proportional to the expectation value of the interaction Hamiltonian, and the dipolar contribution to the observed resonance frequency is:

$$\Delta\vartheta_D = -\frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi\, r^3}\left(\frac{3\cos^2\theta - 1}{2}\right) \tag{1.27}$$

Residual dipolar coupling

Residual Dipolar Coupling (RDC) leverages dipolar coupling distance orientation in solution NMR by exploiting partial alignment of molecules in solution. This alignment preserves dipolar interactions that are otherwise averaged out under fully isotropic conditions. Weak alignment media, including liquid crystals, bicelles, or stretched gels, can induce this effect. The RDC value is proportional to the degree of molecular alignment and follows:

$$\mathbf{D}_{\text{ij}} = \frac{\mu_0 \hbar \gamma_I \gamma_S}{4\pi\, r^3}\left(\frac{3\cos^2\theta - 1}{2}\right) S \tag{1.28}$$

, where S is the order parameter describing the level of alignment that ranges from 0 for isotropic solutions to 1 for full alignment. Molecular alignment in solution induces weak anisotropic interactions that influence bond vectors, causing deviations from isotropic behavior. These anisotropic effects alter the splitting of non-decoupled chemical shift evolution, resulting in small but detectable differences within the multiplet structures of NMR spectra. Quantifying the magnitude of these splittings and mapping them onto the protein sequence enables estimation of the overall molecular shape and the relative orientation of protein domains.

Nuclear Overhauser effect

The Nuclear Overhauser Effect (NOE) is a consequence of the dipole-dipole coupling and is a key observable for determining macromolecular structure by solution NMR spectroscopy. It reflects the transfer of nuclear spin polarization between two dipole-dipole coupled nuclei. NOE-induced intensity changes occur when selective irradiation of an NMR-active nucleus perturbs the nuclear spin state populations, redistributing spin polarization through dipolar relaxation pathways.

The NOE is governed by double-quantum and zero-quantum mechanisms, where energy is exchanged between nuclear spins and their local environment. In a steady-state NOE experiment, irradiating nucleus I at its resonance frequency disturbs the Boltzmann equilibrium and equalizes its spin state populations. Through dipolar coupling, this perturbed population affects nearby nucleus S, resulting in spin polarization transfer. Fluctuating magnetic fields from molecular motion mediate this interaction.

Dipolar coupling enables the system to restore equilibrium through all permitted relaxation processes. The total Hamiltonian describing the interaction between two coupled nuclear spins I and S is constructed from the Zeeman interactions of the individual nuclei and the dipolar interaction:

$$\hat{\mathcal{H}}_{NOE} = \omega_I \hat{I}_z + \omega_S \hat{S}_z + \sigma \left(\hat{I}_z \hat{S}_z \right) \tag{1.29}$$

, where ω_I, ω_S are the Larmor frequencies of spins I and S, and σ represents the cross-relaxation rate owing to molecular motion. In NOE, cross-relaxation is a key process which allows energy exchange and magnetization transfer between spins. This is governed by the rate equations of relaxation:

$$\begin{aligned} \frac{dI_z}{dt} &= -R_{1I} I_z - \sigma I_z S_z \\ \frac{dS_z}{dt} &= -R_{1S} S_z - \sigma I_z S_z \end{aligned} \tag{1.30}$$

, where R_{1I}, R_{1S} are the longitudinal relaxation rates (T_1^{-1}). The importance of cross-relaxation in NOE is based on the fact that normal spin-lattice relaxation does not induce magnetization transfer, leaving cross-relaxation as the primary alternative pathway. Dipolar interactions can induce cross-relaxation only if the local field fluctuates at a rate similar to the transition frequency. In small, rapidly tumbling molecules, this results in intensity enhancement (positive NOE). In contrast, proteins and other large biomolecules, which tumble more slowly, exhibit intensity

reduction (negative NOE). The NOE enhancement factor is determined by the ratio of the cross-relaxation rate to the total relaxation rate of spin I:

$$\eta = \frac{\gamma_S}{\gamma_I} \frac{\omega_2 - \omega_0}{2\omega_1 + \omega_2 + \omega_0} = \frac{I - I_0}{I_0} \tag{1.31}$$

, where the term $\omega_1+\omega_2+\omega_0$ describes the total relaxation rate of dipolar spin-lattice relaxation, ω_2 - ω_0 is the cross-relaxation rate, I_0 – intensity of spin I in the absence of saturation, I in the presence of saturation. NOE provides interatomic distance constraints that help build 3D structures, they are used for distinguishing cis-trans isomers, NOE intensity changes help characterize molecular motions on ps-ns timescales and are used for understanding protein folding, domain movements, and allosteric regulation.

1.4 Protein dynamics

Molecules are objects full of internal dynamics. Henzler-Wildman and Kern defined protein dynamics as any time-dependent change in atomic coordinates that includes both equilibrium fluctuations and non-equilibrium effects [3]. It describes the structural fluctuations of proteins on multiple timescales, ranging from rapid bond vibrations to slow conformational rearrangements. Protein dynamics is characterized by the timescale, amplitude, and directionality of the fluctuations.

1.4.1 Motional processes

All nuclei oscillate rapidly around their mean positions, a motion known as vibration, which typically occurs on the picosecond timescale. When this motion involves the making and breaking of chemical bonds, it is referred to as chemical exchange. This term also applies when a molecule's conformation changes around a relatively rigid structure, such as a double bond or similar conjugated system. Chemical exchange processes span a wide range of timescales, from nanoseconds to several seconds or longer. In liquids, molecules undergo constant random rotations, which are detectable by NMR due to changes in anisotropic spin interactions, such as CSA and direct dipole-dipole couplings. The timescale of molecular rotation in liquids depends on molecular size, shape, solvent viscosity, and temperature. Small molecules typically rotate on the picosecond timescale, while large molecules, such as proteins, rotate on the nanosecond timescale. Another type of motion in solution is translation, or the movement of the molecule's mass center through space. Random, uncoordinated translational motion is called diffusion, while concerted, directed motion is referred to as flow.

1.4.2 Timescales of protein dynamics

The timescale for protein dynamics is defined by the exchange constant (k_{ex}) relative to the chemical shift timescale ($\Delta\omega$, which represents the difference in chemical shift between interconverting species). The exchange constant, defined as $k_{\text{ex}} = k_{AB} + k_{BA}$, quantifies the frequency of transitions between two states, A and B, per unit time. This parameter is directly related to the equilibrium constant K_{eq}, which specifies the ratio of populations between the two states:

$$K_{\text{eq}} = \frac{[B]}{[A]} = \frac{k_{AB}}{k_{BA}} \tag{1.32}$$

Interconversion between states A and B can occur across a broad range of timescales, from femtoseconds to seconds or longer. Identical K_{eq} values may be observed for different dynamic regimes, provided that the ratio of k_{AB} and k_{BA} remains constant. Three distinct exchange regimes can be defined: slow ($k_{\text{ex}} \ll |\omega_\Delta|$), intermediate ($k_{\text{ex}} \approx |\omega_\Delta|$) and fast ($k_{\text{ex}} \gg |\omega_\Delta|$) (Table 1.1, Fig. 1.2). The parameter ω_Δ scales linearly with the $\mathbf{B}_0$ field strength, so increasing $\mathbf{B}_0$ decreases the relative timescale of exchange for a fixed k_{ex} compared to an increased $|\omega_\Delta|$). Solution NMR detects and quantifies only those internal motions that are faster than the molecule's overall rotational tumbling, as slower motions are averaged out by global rotation [4].

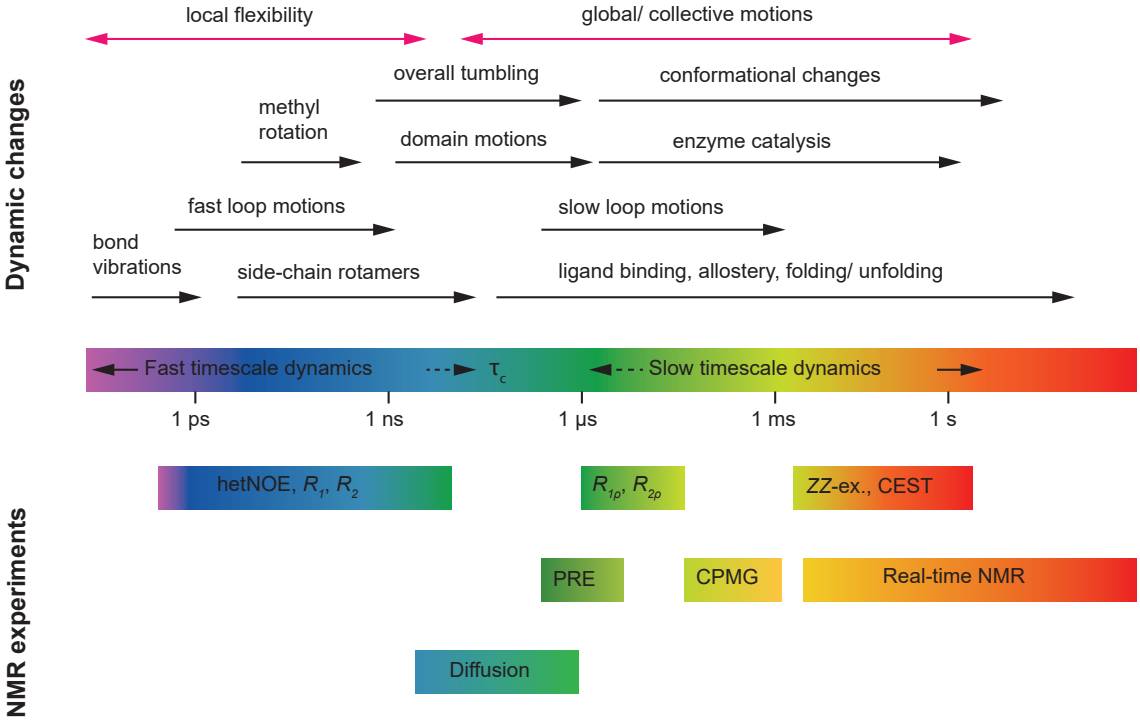


Figure 1.2: Overview of the protein dynamics, timescale and available NMR methods to obtain dynamic information on a biomolecule. Adapted from [5].

Fast regime motions (ps-ns) ($k_{\text{ex}} \gg |\omega_\Delta|$): These include bond vibrations and side chain rotations in structurally similar states that are separated by energy barriers of less than $1kT$ [3]. A single signal is observed, corresponding to a population-weighted chemical shift $\omega_\Delta^{\text{obs}} = p_A\,\omega_A + p_B\,\omega_B$, due to rapid $A \leftrightarrow B$ interconversion and consequent averaging during the detection period. Fast timescale dynamics is usually characterized by amplitude (order parameter S^2) and the timescale of bond fluctuations (internal correlation time τ_c), and are commonly measured for backbone amide bonds and side chain methyl groups [6].

Slow regime motions (ms-s) ($k_{\text{ex}} \ll |\omega_\Delta|$): In the slow exchange regime, signals from both states are observed, each reflecting distinct chemical shifts, intensities, and linewidths. The relatively long lifetimes of each state enable experimental trapping or direct observation, with each state appearing as a distinct peak in the

spectrum and permitting direct characterization. In this regime, the intensity of each peak directly reports on the population of that species [6]. Motions in the slow exchange regime involve rare events, such as fluctuations between kinetically distinct states separated by energy barriers of several kT [3]. Slow timescale dynamics have attracted significant attention recently, as many biological processes, including enzyme catalysis, signal transduction, protein–protein interactions, occur on this timescale [7][8][9].

Intermediate regime motions (μs-ms) ($k_{ex} \approx |\omega_\Delta|$): In intermediate exchange, a single signal is observed at a chemical shift between ω_A and ω_B. The linewidth of this resonance is broadened by an amount R_{ex}, which contributes to the overall measured transverse relaxation rate (R_2^{eff}) as a result of interference from $A \leftrightarrow B$ interconversion during the detection period. In many cases, this broadening prevents peak detection. In this regime, NMR lineshapes are particularly sensitive to the chemical exchange process [10]. Motions occuring in this regime include large-scale rearrangements, such as domain movements, partial unfolding, or transitions between distinct conformational states (e.g. conformational exchange) [11][12]. These motions are frequently functionally significant and may be initiated or suppressed by ligand binding or post-translational modifications, which alter state populations, interconversion barriers, and local interaction networks such as electrostatics, hydrogen bonding, and steric packing [13][14].

Table 1.1: Characteristics of motion regimes

Exchange regime	**Condition**	**Chemical shift**	R_2
Fast exchange	$k_{ex} \gg \lvert\omega_\Delta\rvert$	$\omega_\Delta^{eff} = P_A\omega_A^{int} + P_B\omega_B^{int}$	$R_2^{eff} = P_A R_{2A} + P_B R_{2B}$
Intermediate exchange	$k_{ex} \approx \lvert\omega_\Delta\rvert$	*undefined (peaks coalescence)*	$R_2^{eff} = R_2^0 + R_{ex}$
Slow exchange	$k_{ex} \ll \lvert\omega_\Delta\rvert$	two separate peaks at ω_A^{int} and ω_B^{int}	$R_2^{eff} = R_2^0 + k_{ex}$

In Relaxation Dispersion (RD) experiments, the bulk magnetization vector $\mathbf{M}_0$ resulting from the application of $\mathbf{B}_0$ field is forced to precess by application of a transverse magnetic field $\mathbf{B}_1$. In such cases, chemical exchange will modulate the behavior of magnetization so that the observed relaxation rate R_2^{obs} is comprised of the transverse relaxation rate R_2^0 and an exchange contribution R_{ex} that arises due to intrinsic dynamic properties of atoms.

In practice, to quantitatively assess the exchange rate $k_{\mathbf{ex}}$, populations of exchanging states p_A and p_B, chemical shift differences $\Delta\omega$, and kinetics of rare or invisible states, dispersion-based methods (CPMG and rotating-frame $R_{1\rho}$) are used [15][16][17][18]). In Carr-Purcell-Meiboom-Gill (CPMG) experiments, the contribution R_{ex} to R_2^{obs} is progressively quenched by modulating the CPMG frequency, and the resulting relaxation dispersion curves are modeled using the Bloch-McConnell theory to obtain the chemical shift difference among states, the fractional population of each state, and the rates of exchange among states. Alternatively, $R_{1\rho}$ spin-lock experiments can be used, that offer some advantages relative to CPMG experiments, such as enhanced sensitivity to faster microsecond timescale motions, and ability to robustly determine a complete set of exchange parameters using experiments performed at a single static magnetic field [19][20][19][20]. The CPMG relaxation dispersion experiment is sensitive to exchange occurring in the range from 0.5 ms to 10 ms [10][21][19][22], while $R_{1\rho}$ spin-lock experiment is sensitive to exchange occurring in the range from 0.1 ms to 10 ms [23]. An appropriate ratio of free to bound species is necessary to generate observable RD profiles. In the intermediate regime, chemical exchange-associated line-broadening limits the fraction of protein in the ligand-bound state. Typically, ligand concentrations of 3-25% are used for high-resolution RD experiments [24][25][26].

In addition to relaxation dispersion, several other NMR experiments can indirectly inform on the motional regime of interactions, such as ligand titration. Titration experiments depend on the appropriate choice of protein and ligand concentrations, buffer composition, temperature, and sampling scarcity. Since the observed chemical shift in NMR titration series is a weighted average of the shifts of the free and bound states, it follows a binding isotherm:

$$\Delta\omega^{\text{obs}} = \frac{\Delta\omega^{\text{n}}[L]}{K_{\text{d}} + [L]} \tag{1.33}$$

where $\Delta\omega^{\text{obs}} = |\omega^{\text{eff}} - \omega^{\text{P}}|$ is the observed chemical shift perturbation, $\Delta\omega^{\text{n}} = |\omega^{\text{PL}} - \omega^{\text{P}}|$ is the maximum possible shift perturbation (e.g., fully bound state), $[L]$ is the ligand concentration, and $[P]$ is the protein concentration. The ligand concentration $[L]$ should ideally span $0.1K_{\text{d}}$ to $10K_{\text{d}}$ range; if K_{d} is unknown or poorly estimated, the titration may miss the transition region (where the population of free protein p_{f} is equal to the population of the bound protein p_{b}), resulting in poorly constrained or inaccurate fits.

1.5 ^{1}H-detected MAS NMR

Traditional applications of NMR in structural biology have primarily focused on analyzing proteins and other biomolecules in solution. Solution NMR techniques are particularly effective at probing the structure, dynamics, and interactions of soluble protein monomers and multimers. These methods rely on the rapid tumbling of molecules to suppress or reduce anisotropic and inter-atomic NMR interactions, resulting in narrow spectral lines and enabling multidimensional spectroscopy. However, the necessity for rapid molecular tumbling imposes an upper size limit on target proteins. Larger proteins, vesicle-associated proteins, oligomeric complexes, large assemblies, and aggregates exhibit slow or negligible tumbling. Such samples, characterized by short coherence lifetimes that prohibit detailed characterization by solution NMR, are the primary focus of modern biomolecular Solid-state NMR (ssNMR) [27].

1.5.1 History of biomolecular solid-state NMR

Historically, high-resolution NMR developed first in solution, where proton detection was favored because the high gyromagnetic ratio of ^{1}H provides maximal sensitivity and thus the best Signal-to-Noise Ratio (SNR) compared to ^{13}C and ^{15}N detection. ssNMR, in contrast, initially emerged as a ^{13}C-detected technique for powdered samples, where strong homonuclear ^{1}H-^{1}H dipolar couplings, CSA and quadrupolar interactions severely broaden proton lines and limit resolution. NMR was first observed in solids and liquids in 1946: in solid paraffin [28], and in liquid water [2]. These early experiments quickly revealed that dipolar couplings in solids scale as $r^{-3}\,(3\cos^2\theta - 1)$ [29]. Soon after that it was shown that the resulting linewidths encode inter-nuclear distances and orientations [30]. The technical breakthrough that made high-resolution ssNMR possible was the introduction of Magic Angle Spinning (MAS) $\theta = \arccos\left(\sqrt{\frac{1}{3}}\right) \approx 54.74°$, and the recognition that rapid spinning averages anisotropic interactions into a narrow central band accompanied by spinning sidebands [31][32][33]. FT-NMR, high-field magnets, average Hamiltonian theory and related formalisms [34][31], and improved rotor designs [35] transformed both solution- and ssNMR into quantitative spectroscopies. In biomolecular systems, Cross-Polarization (CP) allowed efficient transfer of polarization from abundant protons to dilute nuclei such as ^{13}C and ^{15}N and thus boosted signals from low-abundance sites [36][37]. Multipulse MAS sequences [38] and early 2D experiments [39] extended spectral dispersion and enabled structural studies of complex systems such as membrane proteins [40]. In the 1990s, recoupling techniques including TEDOR and REDOR, provided accurate distance restraints and side chain conformations in membrane-embedded and aggregated proteins [41][42]. 3D pulse sequences developed in solution NMR [43] were rapidly adapted to MAS [44][45][46][47], giving ssNMR the same kind of multidimensional spectral resolution that had driven the expansion of solution protein NMR. To address the intrinsic low sensitivity and long measurement times of multidimensional experiments, Dynamic Nuclear Polarization (DNP) was introduced into MAS NMR, building on Overhauser's concept and later electron-nuclear polarization-transfer studies [48][49][50]: electron spin polarization from exogenous radicals is transferred to nuclei, yielding

order-of-magnitude signal enhancements and making many previously impractical experiments feasible. In parallel, microcoil and cryogenic probe technologies further increased sensitivity [51], while advanced recoupling strategies continued to refine coherence transfer and distance measurements [52][53]. Pulse-sequence design itself has evolved from hand-crafted schemes based on product operator analysis, Floquet theory and average Hamiltonian theory to numerically optimised experiments using Optimal Control (OC) theory [54][55]. OC treats the RF waveform as a high-dimensional control problem and can search thousands of degrees of freedom to maximise transfer efficiency or robustness, surpassing traditional low-parameter optimisations and often revealing non-intuitive pulse solutions. OC algorithms have been implemented in simulation packages such as SIMPSON [56] and the SPINACH library [57], and applied to a range of ssNMR experiments: improved CP at 1.2 GHz [58], OC pulses for N–Cα and N–CO transfers in backbone assignment [59][60], and Cα-CO mixing sequences [54], which are now curated in dedicated pulse-sequence libraries. Miniaturisation and optimisation of MAS hardware, in particular rotor and stator ceramics, have paralleled these developments; current 0.4 mm rotors require on the order of 0.16 μL of sample and allow MAS frequencies up to 160 kHz. Together, these experimental and theoretical advances have established modern ssNMR as a high-resolution, high-sensitivity technique capable of probing both structure and dynamics of large protein assemblies, membrane proteins and disordered systems in environments that closely resemble their native states.

1.5.2 Key techniques of biomolecular ssNMR

The Hartmann-Hahn condition is a fundamental principle in ssNMR and is essential for the success of CP experiments. This condition requires that the RF fields applied to two different types of nuclei, such as ^{1}H and ^{13}C, are matched to enable efficient magnetization transfer between them [61]. In solution and in static solids, the condition is expressed as follows:

$$\gamma_I \mathbf{B_{1,I}} = \gamma_S \mathbf{B_{1,S}} \tag{1.34}$$

, where γ_I and γ_S are the gyromagnetic ratios of the two nuclei (^{1}H and ^{13}C/^{15}N, correspondingly), and $\mathbf{B_{1,I}}$ and $\mathbf{B_{1,S}}$ are the strengths of the RF fields applied to each nucleus. When this condition is met, spin-locking RF fields establish resonance between the two nuclei, facilitating efficient dipolar-driven magnetization transfer. Under MAS, the condition is modulated by the spinnin frequency in various ways. This process constitutes the foundation of CP and is critical for enhancing the signals of nuclei such as ^{13}C or ^{15}N, which are less abundant or exhibit lower sensitivity. CP is particularly beneficial for the study of biomolecules, as it substantially reduces the acquisition time required to obtain high-quality spectra. Furthermore, CP frequently increases signal intensity for nuclei present at low concentrations, including isotopically labeled sites in proteins or sparsely distributed functional groups in complex systems.

CP is the most frequently used polarization transfer sequence in ssNMR, and is particularly efficient in rigid systems with exchange rate constants of 10^{-5} s^{-1}. Alternatively, polarization transfer can be mediated by scalar coupling, as implemented in the Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) experiment

[62][63]. INEPT-based sequences are more efficient in flexible systems with exchange rate constants around 10^{-7} s^{-1} [64].

Recoupling

Recoupling methods are essential for reintroducing specific interactions that are otherwise averaged out by MAS. Recoupling can be broadly categorized according to the type of interaction being reintroduced, such as dipolar recoupling, and CSA recoupling. A typical recoupling block in a pulse sequence serves two primary functions: (a) to cancel the effect of MAS by applying a train of rotor-synchronized pulses, and (b) to suppress other interactions through phase cycling, composite pulses, or related strategies that prevent unwanted interactions from contaminating the spectrum. The design of recoupling pulse sequences commonly employs symmetry theory, Floquet theory, average Hamiltonian theory, and group theory to address unwanted Hamiltonian terms while retaining or reintroducing the desired interaction.

Recoupling techniques are particularly valuable in structural studies of biomolecules, where distance constraints derived from these methods can inform molecular modeling. Dipolar recoupling methods, such as REDOR and DRAWS, are primarily used to measure internuclear distances [65][66]. CSA recoupling techniques yield information about electronic environments and molecular orientations, making them particularly valuable for understanding ligand binding in active sites or characterizing conformational changes in proteins [67][68].

Decoupling

Decoupling plays a critical role in MAS NMR by suppressing undesired interactions, such as heteronuclear and homonuclear dipolar couplings, which otherwise cause line broadening and diminished resolution. High-power decoupling schemes like TPPM and SPINAL-64 are commonly employed to achieve narrow linewidths [69][70]. The physical basis of decoupling involves the application of a continuous or pulsed RF field during acquisition, which causes unwanted interactions like ^{1}H-^{1}H or ^{1}H-X dipolar coupling to be effectively eliminated. In biomolecular studies, heteronuclear decoupling is crucial for detecting signals from nuclei such as ^{13}C and ^{15}N in the presence of abundant ^{1}H. Decoupling enhances spectral resolution and also aids in unambiguous signal assignment.

Relaxation effects in ssNMR

In ssNMR, just as in solution NMR, spins relax to thermal equilibrium through several mechanisms. The main relaxation pathways include T_1 and T_2 relaxation. Additionally, dephasing is strongly affected by the incomplete averaging of anisotropic interactions. Similar to solution NMR, T_1 in ssNMR measures how rapidly spins return to their equilibrium magnetization along the $Z-$axis. The dense network of protons can shorten T_1 by providing many dipolar pathways for energy exchange. Shorter T_1 values are advantageous for faster data collection, but they can become problematic if the magnetization does not fully recover between transients. This issue is especially relevant in multi-dimensional experiments, where incomplete recovery may reduce signal in intermediate coherence transfers. T_2 reflects the decay

of phase coherence among spins in the $XY-$plane and is a critical determinant of coherence lifetime. In rigid solids, efficient dipolar interactions, particularly ^{1}H-^{1}H and ^{1}H-X, dominate transverse relaxation, resulting in short T_2 values that constrain the duration of evolution periods in multi-dimensional NMR experiments. However, coherence lifetimes have been shown to increase approximately linearly with both MAS frequency and static magnetic field $\mathbf{B}_0$ strength [71]. Sample modifications are implemented to mitigate the adverse effects of spin relaxation. Most approaches involve altering the isotopic composition of the protein (protein labeling) to minimize T_2 relaxation. Deuteration of carbon moieties significantly reduces dipolar relaxation by a factor of more than an order of magnitude for autocorrelated mechanisms [72]. One important method to modify relaxation properties of a protein is Paramagnetic Relaxation Enhancement (PRE), or doping by introducing metal ions as chelates to accelerate T_1 relaxation while preserving T_2 relaxation [73]. The paramagnetic properties of metal ions are dictated by their oxidation state, coordination sphere, and electron spin states, so different ligands can fine-tune these properties. Chelators such as EDTA and DO2A [74], DTPA [75], TAHA [76], DOTA [77][78], and derivatives [79][80] are widely used to develop tagging strategies that generate different paramagnetic effects using the desired metal. All of these complexes rely on unpaired electrons to provide additional relaxation pathways, but they differ in how intensely they broaden resonances: Mn^{2+}- or Gd^{3+}-based dopants often cause more line broadening than Ni^{2+} or Cu^{2+} chelates, due to slower electron-spin relaxation. The most common doping agents are Gd^{3+} complexes and Cu-Ethylenediaminetetraacetic Acid (EDTA) [81]. Combination of doping with protein perdeuteration allows to further improve relaxation properties of a crystalline sample by shortening T_1 and extending T_2, which increases the SNR per experimental time (i.e. sensitivity).

Relaxation pathways T_1, T_2, and $T_{1\rho}$ must be carefully controlled in ssNMR experiments to balance sensitivity, resolution, and experimental speed. Key variables such as proton density, MAS rate, decoupling power, deuteration level, and paramagnetic doping can be systematically adjusted to enhance data quality for backbone assignments and other structural studies.

1.5.3 Multidimensional ssNMR experiments

The prerequisite to structural or functional studies of proteins using NMR is the assignment of NMR resonances to individual atoms within the protein. Resonance assignment of protein NMR signals in solution generally follows a standard protocol: (a) clustering chemical shifts of nuclei of interest (^{1}H, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO) into spin systems that represent candidate residues; (b) aligning spin systems sequentially by using inter- and intra-residual correlations that show which spins interact with neighboring residues along the protein chain; (c) comparing chemical shift data for each spin system to shift signatures of specific amino acid types (Gly, Ala, Ser, Thr), and linking them to cover a polypeptide chain, ultimately yielding residue-by-residue assignment of all observed resonances.

In the general assignment procedure, H–N correlations play a central role because all residues, with the exception of prolines and the N-terminal residue, possess a single pair of amide proton and nitrogen. Their resulting 2D-correlation map serves as a starting point for the assignment and is used for monitoring the progress [82]. Various 3D, 4D, or 5D spectra connect each H-N pair to the carbon and proton

resonances of both their own and the preceding residue, thereby enabling sequential linkage of H-N pairs. $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts provide information about the residue type and, to some extent, secondary structure [83].

A standard ^{1}H-detected 2D pulse program, such as hNH, typically begins with a CP block from ^{1}H to ^{15}N to establish initial magnetization on the nitrogen channel, followed by t_1 evolution on the ^{15}N nucleus. Traditionally, this process is performed under high-power ^{1}H-decoupling to suppress strong homonuclear dipolar interactions. Following evolution, magnetization is transferred back to ^{1}H via a second CP step, enabling detection on the proton channel under heteronuclear decoupling. The sequence incorporates rotor-synchronized delays to manage coherence transfer under MAS and often incorporates Z-filters or phase cycling schemes to eliminate residual transverse magnetization. Compared to its solution-state counterpart Heteronuclear Single Quantum Coherence (HSQC), the HSQC, the hNH sequence in solids avoids long INEPT delays due to short T_2 values and instead relies on CP-based transfer, which is more robust under fast spinning and strong dipolar couplings.

After recording a fingerprint hNH spectrum, a third dimension can be introduced, e.g. to encode the carbonyl chemical shift(s) from from residue i, $i-1$, or $i+1$. Among the available 3D experiments, the 3D hCONH is most sensitive. The 3D hCONH pulse sequence correlates the carbonyl (^{13}CO) chemical shift of residue $i-1$ with the amide nitrogen (^{15}N) and amide proton (^{1}H) of residue i, forming a key part of sequential assignment strategies in ^{1}H-detected ssNMR. Another frequently employed 3D experiment for backbone resonance assignment is 3D hCANH, which correlates the ^{1}H and ^{15}N resonances of residue i with $^{13}C_{\alpha}$ resonance from the same residue. It essentially differs from 3D hCONH by the frequency of the ^{13}C-channel carrier and power and duration of carbon pulses. Standard 3D experiments can, in turn, serve as source experiments for 4D pulse sequences, such as 4D hCACONH, 4D hCOCANH, 4D hCACBcaNH, and 4D HNcocaNH [84][85]. The use of 4D experiments for backbone assignment in large proteins is necessitated by significant peak overlap in ssNMR spectra. The 4D hCOCANH and 4D hCACONH experiments correlate the i and $i+1$ amide shifts based on shared $^{13}C_{\alpha}$ and ^{13}CO resonances. The 4D hCOCANH is analogous to a 3D HNcaCO but incorporates an additional $^{13}C_{\alpha}$ evolution period, providing a comprehensive set of ^{1}H, ^{15}N, $^{13}C_{\alpha}$, and ^{13}CO chemical shifts for each i residue. The 4D hCACBcaNH experiment is a 4D-derivative of a 3D hCANH with additional $^{13}C_{\alpha}$-$^{13}C_{\beta}$ INEPT transfer and evolution on $^{13}C_{\beta}$. This experiment is analogous to a solution NMR 3D HNCACB experiment with an extra carbon dimension. The 4D hCACBcaNH ssNMR pulse sequence is designed to trace magnetization from the amide proton ^{1}H through the backbone and side-chain carbons ($^{13}C_{\alpha}$ and $^{13}C_{\beta}$), returning to the amide proton for detection, thereby enabling sequential assignment in solids.

The choice of assignment method in ssNMR depends largely on a variety of factors including available hardware, sample preparation and isotopic labeling scheme, and sample-dependent linewidths and spectral quality.

Chapter 2

Tackling the H/D exchange problem

This chapter is based on work published in [86] and is reproduced here with minor modifications.

All work on SH3 domain of chicken α-spectrin was performed by Dr. Sara Medina Gomez and Dr. Alexander Klein.

During the course of this work, an independent study by Paul Schanda group reported a closely related approach to application of algal cell lysates to boost protein perdeuteration for MAS NMR [87]. That study appeared two months after our own publication and provides an important external validation of the concepts developed here.

2.1 Introduction

Protein perdeuteration has evolved as a fundamental technique in current NMR-based structural biology [88]. Initially introduced for solution NMR, perdeuteration has significantly improved the relaxation properties of ^{1}H spins and heteronuclei. Fast MAS NMR has emerged as a potent alternative to solution NMR-based characterization of protein structure, dynamics, and interaction [89]. Compared to traditional ^{13}C-detected experiments, ^{1}H detection has become very popular due to its inherently higher sensitivity. Extensive deuteration of proteins with observation of only a few remaining protons on amide-, methyl- or other side-chain sites, has proven to be a viable way to obtain high resolution and sensitivity [90][91][92][86].

However, cell cultivation in D_2O imposes several drawbacks, such as high costs, low growth rate [93], and low yields [94]. It is also incompatible with insect, mammal, and plant cell cultures [95] [96]. The mechanistic basis of reduced growth in D_2O are not known, however the deuterium kinetic isotope effect likely contributes. Various strategies have been devised to mitigate poor growth in D_2O. These include gradual adaptation of bacteria to increasing concentrations of D_2O in media, directed evolution of *E.coli* for rapid growth in deuterated media [97], and enrichment of perdeuterated culture with nutrients [98]. Deuterium adaptation is the most common of them all and is usually performed as a gradual transition of bacterial culture from 100% LB/H_2O media to 100% M9/D_2O media over the span of 2-4 days.

Another limitation arising due to usage of D_2O as a solvent is incomplete H/D

exchange. Upon bacterial cultivation in D_2O, all hydrogen atoms are replaced with ^2H atoms, which effectively renders them invisible for standard ^1H-detected NMR. On one hand, perdeuteration of aliphatic side chains narrows proton linewidths due to improved relaxation properties. On another hand, amide protons must be reintroduced by H/D back exchange during the course of protein purification to be observable. For large proteins with a substantial hydrophobic core, H/D exchange will likely not reach 100% due to lack of interaction with the solvent, and hydrophobic core amides will remain invisible [99]. A common workaround is unfolding using chemical denaturants followed by renaturation in H_2O. This is not always feasible, as for many proteins refolding can reduce catalytic activity, promote protein precipitation, and introduce conformational heterogeneity and poor reproducibility [100][101].

Several approaches can be used to partially resolve the issue: (a) avoid perdeuteration in the first place and instead increase MAS speed to counteract the line broadening [102][103], (b) Inverse Fractional Deuteration (iFD) approach [104], (c) cell-free synthesis in H_2O [105][106], and (d) usage of isotopically labeled perdeuterated amino acid mixes [107]. All methods have their strengths and weaknesses. For instance, fast and ultra-fast MAS with fully protonated protein samples requires smaller rotors (0.4-0.7 mm) and, therefore, provide less sensitivity due to the lower amount of protein used. Even at 160 kHz MAS, fully protonated samples tend to have larger amide ^1H resonance linewidths and shorter coherence lifetime compared to fully deuterated proteins at 60 kHz MAS [102][103]. In practice, ultra-fast MAS combined with multidimensional experiments and Non-Uniform Sampling (NUS) schemes can result in excellent sensitivity and resolution. However, feasibility depends strongly on molecular weight and relaxation properties of a protein in question.

The iFD approach has been used to quantitatively incorporate amide protons while maintaining partial deuteration of the side chains in transmembrane proteins [104]. This technique involves protonated water while exclusively deuterating the carbon source (usually glucose) of a minimal medium. This results in low (10-40%) and random deuterium incorporation in side-chains but in 100% ^1H occupancy at $^1H_\alpha$ sites, including those in exchange-protected regions. Cell-free synthesis has the advantage that various NMR isotope labeling schemes, including amino-acid-selective labeling, can be easily implemented. Importantly, deuteration in combination with complete amide protonation can be achieved directly during synthesis, avoiding a denaturation and refolding step [105][106].

Finally, highly deuterated proteins with full protonation at all amide sites can be prepared by cultivating *E.coli* in deuterated algal or bacterial lysates in 100% H_2O. In solution NMR, this procedure yields approximately 80% deuteration at aliphatic sites while the deuteration at $^{13}C_\alpha$ sites is significantly lower due to the incorporation of water protons during biosynthesis of amino acids depleted during bacterial growth. Löhr et al. implemented this approach to obtain nearly full backbone assignments of a 30 kDa DFPase enzyme, which provided only fragmentary and ambiguous sequential assignments with a standard perdeuteration labeling scheme [107]. Tekwani Movellan et al. modified this approach by converting commercially available $^2H/^{15}N/^{13}C$-labeled amino acid mixture into a mixture of keto acids using L-amino acid oxidase from a crude snake venom to obtain more quantitative $^1H_\alpha$ protonation [108]. They showed that replacement of perdeuterated ^{13}C -labeled glu-

cose and ^{15}N-labeled ammonium chloride with the mixture of α-keto acids results in up to 100% $^{1}H_{\alpha}$ incorporation, which allows one to obtain backbone resonance assignments at below 60 kHz MAS. Application of deuterated amino acid and keto acid mixtures in H_2O has several advantages, such as (a) adaptation of the cells to D_2O is not necessary, which makes sample preparation faster and easier; (b) it gives higher protein yields compared to deuterated expression in M9 media; (c) costs are similar to production of deuterated proteins; (d) it solves the problem of H/D exchange and results in hydrophobic core residues becoming visible for ^{1}H-detected NMR.

2.2 Aim of the project

In this project, we aimed to: (a) evaluate whether a purely cell-extract-based labeling protocol can match or surpass the protein yields of standard M9/D_2O culture, (b) determine which algal or bacterial extracts provide optimal proton incorporation, and (c) quantify the number of previously silent residues that reappear in ^{1}H-detected 3D/4D MAS spectra. Achieving these objectives would benefit a wide range of enzymes and membrane complexes in which solvent-shielded amides encode allosteric communication but remain undetected.

2.3 Case study: Tryptophan synthase as a benchmark system

Tryptophan synthase (TS, $\alpha\beta\beta\alpha$ heterotetramer, 144 kDa) contains a well-characterised hydrophobic core that is resistant to proton back-exchange. Earlier work on *Salmonella typhimurium* TS from our group achieved only partial backbone coverage under classical perdeuteration [85].

2.4 Materials and Methods

2.4.1 Media screening

Three different cell extracts were tested: ISOGRO®, SILEX®, and BioExpress®. LB and minimal M9 media served as reference. Details on chemical composition, features, and approximate price per 1 L of a labeled bacterial culture can be found in Supplementary Table A.5. To determine which media is the most efficient at boosting TS expression, we utilized *E. coli* CB149 cells harboring pEBA-TrpA-TrpB vector encoding both subunits of TS (TrpA encoding for α, TrpB encoding for β) from *S. typhimurium*. Freshly transformed cells were inoculated into 5 ml of each of five media tested, and cultures were supplemented with carbenicillin to a final concentration 100 μg/ml. Optical density of cultures was tested every hour, and the media that provided the fastest growth was considered optimal. Presence of TS protein was then confirmed by SDS-PAGE (see below). Best-performing extract was selected as a compromise between the price and protein yield.

2.4.2 Expression and purification

Detailed buffer compositions can be found in Supplementary Table A.4.

Perdeuterated, u-^{13}C/^{15}N minimal media was prepared according to standard protocols [89] on the basis of 2 g of ^{2}H/^{13}C-labeled glucose and 1 g $^{15}NH_4Cl$. Media for algal amino acid labeling was prepared by dissolving 10 g of ^{2}H/^{15}N/^{13}C-labeled ISOGRO® powder in a small volume of deionized water, followed by addition of 18 ml of 100 g/L K_2HPO_4 solution, 28 ml of 50 g/L KH_2PO_4 solution, 20 ml of 50 g/L $MgSO_4$ solution, and 0.3 ml of 37 g/L $CaCl_2$ solution. Media pH was adjusted to 7.0 with NaOH prior to volume adjustment, supplemented with 100 mg of ampicillin, and filter-sterilized using a 0.2 μm filter. Media for iFD samples prepared according to Medeiros-Silva et al. [104] as an H_2O-based standard M9 minimal media containing $^{15}NH_4Cl$ and ^{2}H/^{13}C-glucose.

SH3. Chicken α-spectrin SH3 domain samples were purified in three labeling schemes: (a) iFD sample, (b) ISOGRO® sample, and (c) ^{1}H/^{15}N/^{13}C-sample from a standard minimal M9 media as a reference for quantification of side chain protonation. In all cases, proteins were expressed in *E. coli* BL21-DE3 cells grown at 37 °C with induction of expression at an OD_{600} of ≈0.8 using 1 mM IPTG. After induction, the cells were incubated overnight at 20°C and harvested after 18 h by centrifugation at 6,000 g at 4°C for 20 min, and pellets were flash-frozen in liquid nitrogen and stored at -80°C. All SH3 protein samples for the solution NMR were purified using established chromatography methods [90]. Briefly, the cell pellet was thawed at 4°C and resuspended in 1:5 (v/v) volume of **buffer A**. Cell suspension was supplemented with 10 μg/mL DNase I and one Complete Protease Inhibitor Cocktail tablet, and incubated at 4°C for 30 min. Cells were then disrupted in a high-pressure homogenizer Emulsiflex C3 in two rounds and centrifuged at 110,000 g for 45 min. The supernatant was separated by decantation and diluted with **buffer A** to approximately 80% of the Column Volume (CV) of the anion exchange column. The solution was injected into the HiPrep QFF 16/10 column with a flow rate of 3 ml/min, followed by the column wash with 2 CV of buffer A. Protein was eluted with a linear gradient of 0-9% of **buffer B** (8 CV). Elution fractions were assessed with SDS-PAGE; the SH3-containing fractions were pooled together, and pH was adjusted to pH 3.5 using 3 M citric acid. Protein was then concentrated to 10-12 mg/ml and further purified on a Superdex 75 column. Purified SH3 protein was then concentrated for solution NMR studies.

TS. TS samples were purified in two labeling schemes: (a) ISOGRO® sample, and (b) ^{2}H/^{15}N/^{13}C-sample from a standard minimal M9 media as a reference. Algal amino acid labeled and perdeuterated, u-^{15}N/^{13}C-labeled TS samples for MAS NMR measurements were expressed using the above-described growth media upon induction via 0.4 mM IPTG at an OD_{600} of 0.6-0.8. TS expression and purification was performed according to the protocol described by Hilario et al. [109]. Briefly, *E.coli* CB149 cells harboring pEBA-TrpA-TrpB construct were inoculated from a Petri dish using a pipette tip, and the culture was left to shake at 37°C at 160 rpm until the OD_{600}=1.2. The culture was then cooled down on the bench for 30 min, after which protein expression was induced by adding 0.4 mM IPTG. The culture was incubated overnight at 30°C upon shaking at 160 rpm. The cells were then harvested by centrifuging at 4,500 g at 4°C for 20 min, the supernatant was removed, and the cells resuspended in cold **buffer T** (5 ml of buffer/g cells). Cells were disrupted by two runs through the high-pressure homogenizer Emulsiflex C3.

The cell lysate was centrifuged at 80,000 g for 30 min at 4°C, the supernatant was collected and the pellets discarded. **Buffer T supplemented with 30% (w/V) PEG8000** was added to the supernatant (0.25 ml/ml of supernatant), followed by addition of 0.5 M spermine (0.0125 ml/ml of supernatant). The sample was then transferred to centrifuge tubes, spun down for 10 min at 80,000 g at 4°C. The supernatant was collected and left at 4°C on a stir plate for 48 h. After 24 h, **buffer T with 30% PEG8000** was added ($V_{30\%\,\text{PEG-8000}} = \frac{V_{\text{supernatant}}}{23}$) and stirred for another 24 h. The protein solution was spun down at 80,000 g for 30 min at 4°C, the pellets were collected and washed in **buffer T containing 30% PEG8000** and 0.5 M spermine (3.2 ml of buffer per gram of cells). After washing, the pellet was resuspended in **buffer B** with pH 9 (4 ml/g of cells). The solution was then stirred at room temperature for 2 h until all protein dissolved and the insoluble fraction becomes white. The solution was then dialyzed twice against 2 L of **buffer B** for 24 h at 4°C. After that, the protein was dialyzed against **buffer P** for 48 h at 4°C, the buffer was changed once after 24 h. The protein precipitate was then collected, dissolved in **buffer B** and dialyzed against **buffer I** three times. The final protein solution was concentrated to 15 mg/ml and used for crystallization. Microcrystalline samples of TS were prepared by diluting the protein solution 1:1 with **buffer C** containing 3 mM N-(4′-trifluoromethoxybenzenesulfonyl)-2-aminoethyl phosphate (F9; an analogue of the natural α-site substrate 3-indole-D-glycerol-3′-phosphate (IGP)) as previously described [92]. Microcrystals were collected and washed with **buffer C**, doped with ≈10 mM Cu-EDTA and packed into a Bruker 1.3 mm MAS rotor.

2.4.3 Protein quality and quantity measurements

SDS-PAGE. Briefly, samples were denatured and reduced by heating in a sample buffer containing Sodium Dodecyl Sulphate (SDS) and a reducing agent (β-mercaptoethanol) for 10 min at 98°C. The proteins were then separated by molecular weight using a 4-14% gradient SDS-polyacrylamide gel electrophoresis in a glycine buffer system. Electrophoresis was performed at a constant voltage of 140 V using a vertical electrophoresis system with a standard Tris-glycine-SDS running buffer. A protein molecular weight marker Color Prestained Protein Standard, Broad Range (10-250 kDa) was run in a separate lane for molecular weight estimation. Following electrophoresis, proteins were visualized by staining the gel with Coomassie Brilliant Blue R-250 stain for 5 min at 95°C, followed by destaining in boiling water until clear bands were visible.

ImageJ Densitometry. Band intensities corresponding to TrpB subunit in lysates produced with various amino acid extracts were quantified in ImageJ (NIH) (https://imagej.net/ij/). For each lane, an identical rectangular Region of Interest (ROI) was placed around the TrpB band, and the integrated density was measured after background subtraction using an adjacent gel region of equal size. TrpB amount in control sample was used as a reference for normalization, corresponding to 100% amount.

2.4.4 NMR data acquisition and processing

For all solution NMR experiments, 2 mM SH3 samples in a 2 mM citrate buffer at pH 3.5 were used. Slight deviations in the protein concentration were factored in

by a constant factor measured from the ^{1}H bulk signal from an HSQC experiment. Solution 2D H(N)(CO)CA spectra using ^{2}H but no ^{1}H decoupling were recorded for the sample based on algal amino acid labeling as well as the iFD sample to assess the amount of $^{1}H_{\alpha}$ protonation. The principles of this strategy are described in more detail in Asami et al.[92]. An additional 2D HNcoCA spectrum with ^{1}H decoupling was recorded for a fully protonated sample for reference and resonance assignment transfer. For the ^{1}H coupled spectra the Bruker pulse sequence hncocagp2h3d was modified by simply removing ^{1}H decoupling pulses. The 2D data sets were recorded using 128 scans and acquisition times of approximately 25 and 90 ms for ^{13}C and ^{1}H, respectively. To quantitatively estimate the amount of protonation in all other aliphatic sites, ^{13}C constant-time HSQC spectra were recorded for all three SH3 samples, (ISOGRO®, iFD, ^{1}H/^{15}N/^{13}C sample), whereas the spectra for the ISOGRO® and iFD samples were acquired using ^{2}H decoupling. All spectra were recorded using 48 scans and the constant-time period set to $2/J_{CC}$, i.e., 26.6 ms. In all cases, the ^{1}H dimension was recorded for 82 ms, whereas approximately 25 ms were used again for the indirect ^{13}C dimension.

^{1}H/^{15}N/^{13}C triple-resonance experiments on microcrystalline TS were performed at 16.4 T (700 MHz ^{1}H Larmor frequency) on a Bruker NEO spectrometer equipped with a triple-resonance 1.3 mm MAS probe spinning at MAS rates of 55.5 kHz. The ^{1}H chemical shifts were referenced with respect to water, and ^{15}N and ^{13}C chemical shifts were referenced indirectly. Each experiment was recorded in two blocks with 2048 points in the direct dimension, 180 points in F2 (^{15}N), and 150 in F3 (^{13}C); number of scans being 8, acquisition time 50 ms in the ^{1}H dimension, 32 ms in the ^{15}N dimension, and 27 ms in the ^{13}C dimension. The total experimental time per experiment amounted to 6 days. Linewidths were measured using apodization with 50 Hz exponential line broadening.

NMR data were processed using TopSpin 4.5.0 (Bruker) (https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html) and analyzed using CCPNMR Analysis v. 3.2.2 (https://ccpn.ac.uk/)[110]. For processing and plotting, custom in-home Python scripts were used.

2.5 Results

2.5.1 Bacterial growth in enriched media

Among the media tested, BioExpress® media supported the fastest biomass accumulation, with growth parameters indistinguishable from the conventional LB media up until mid-linear phase (Fig. 2.1, A). Six hours after inoculation, cultures in LB and BioExpress® media reached OD_{600} = 0.69 and 0.67, respectively. ISOGRO® culture lagged, requiring 10 h longer to reach the induction threshold, while SILEX® media reproducibly exhibited the poorest performance, achieving OD_{600}=0.6 only after 20 h. Despite these differences in growth kinetics between cultures, SDS-PAGE analysis of post-induction lysates and supernatants revealed comparable TS expression levels in BioExpress®, ISOGRO®, and LB preparations. By contrast, the SILEX® culture yielded markedly less target protein which can be explained by the reduced biomass. Quantitative protein-to-cell-mass ratios were not determined, as the goal of this screen was simply to identify the supplement that could replace glucose and $^{15}NH_4Cl$ the most efficiently.

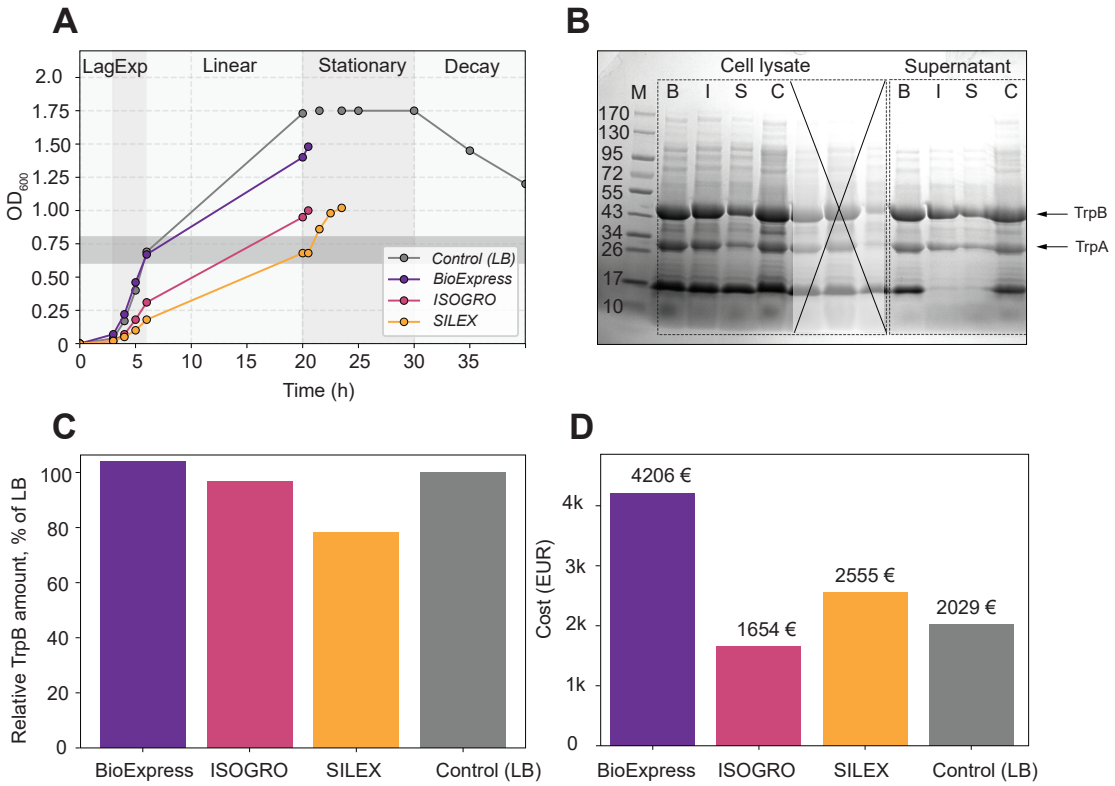


Figure 2.1: **(A)**. Optical density curves of *E.coli* cultures grown with various supplements. **(B)**. SDS-PAGE control of TS expression in 5 ml bacterial cultures. Cell lysate and supernatant fractions have been assessed. TS is comprised of two TrpA (27 kDa) and two TrpB (43 kDa) subunits. (B, BioExpress®; I, ISOGRO®; S, SILEX®; C, Control (LB). **(C)**. Relative amount of TrpB calculated from SDS-PAGE gel TrpB band intensities using ImageJ. TrpB band intensity from control (LB) sample was used as a reference (100%) value. Only lysate TrpB concentrations shown. **(D)**. Calculated price of 1 L of $^{2}H/^{15}N/^{13}C$-culture produced with each of the extracts tested; a standard perdeuterated LB culture is used as a control here.

At first, we selected BioExpress® for scale-up due to its faster growth, but a follow-up cost analysis showed that ISOGRO® powder is cheaper, delivers similar protein yields, and was available at short notice (Supplementary Table A.5); ISOGRO® was therefore chosen for all subsequent preparations. The approach utilized here is a simplistic way to determine a potential cell growth booster which may suffer from the lack of statistical error estimates and absent protein-to-cell ratio calculations.

2.5.2 Heterogeneity of TS crystals

The standard protocol for TS crystallization is based on batch crystallization performed by directly combining protein solution with precipitant in Eppendorf tube [109]. Crystals produced under these conditions can exhibit a range of morphologies and colors, including white powder-like microcrystals and well-defined green, yellow and brown prismatic macrocrystals (Fig. 2.2, B). Such a morphological heterogeneity is undesirable as it may render spectra recorded on different crystalline preparations incomparable (Fig. 2.2 D).

At the NMR level, the initial 2D hNH spectrum obtained from freshly grown TS crystals, exhibited substantial peak offsets compared to the reference data from [85]. At a MAS rate of 55 kHz, the 144 kDa complex produces a crowded central region, yet individual resonances were clearly displaced in both dimensions. We presumed that $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}CO atoms would also have various degree of offset, making direct transfer of existing assignments impractical. To address the observed discrepancy, the crystals were rinsed and resuspended in the TS crystallization buffer with dopant Cu-EDTA buffer used by Klein et al. in 2022 [85]. The use of this 'aged' buffer restored the chemical shifts of certain migrated peaks and, due to the presence of paramagnetic dopant, revealed resonances otherwise invisible. However, line broadening and limited chemical shift dispersion in 3D- and 4D-MAS NMR experiments prevented reliable backbone assignment transfer.

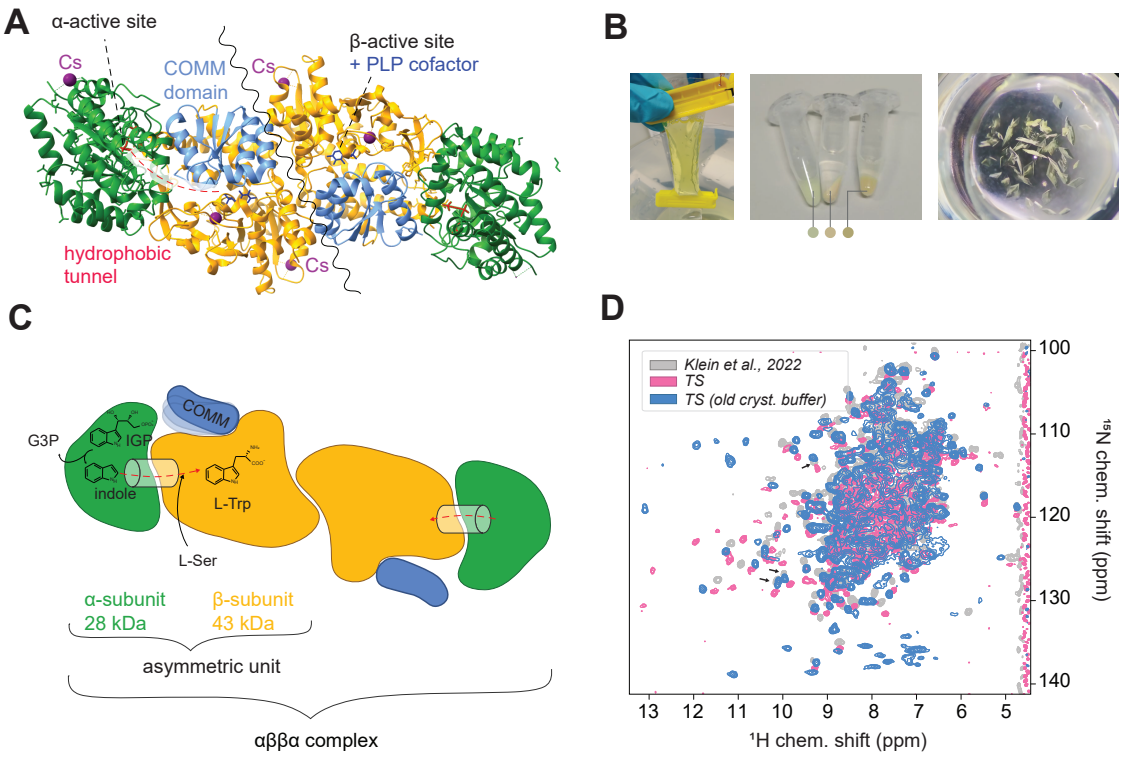


Figure 2.2: **(A)**. Crystal structure of the TS $\alpha\beta\beta\alpha$ tetramer with the COMM domain, hydrophobic tunnel, TrpA active site bound to F9, TrpB PLP site, and catalytic monovalent cations. **(B)**. Schematic of the $\alpha\beta\beta\alpha$ assembly. **(C)**. Photographs: dialysed TS solution (left); diverse crystal shapes and colors from a single batch (centre); microcrystals grown in a sitting-drop manner after 3 days of incubation at 4°C (right). **(D)**. Overlay of hNH spectra: published reference sample (grey) adapted from [85], undoped preparation (pink), and Cu-EDTA-doped preparation (blue), illustrating spectral variability.

2.5.3 Linewidths and peak heights distribution

Overlaying the hCONH strips showed, that all major peaks observed in the perdeuterated sample are retained in the ISOGRO® preparation, and a number of new peaks become detectable in many ^{13}CO planes (Fig. 2.3 C,D, Supplementary Fig. B.1). The median peak height is identical for both datasets (1.3E+09; SNR = 8.8 for ISOGRO®, 7.9 for ^{2}H/^{15}N/^{13}C), indicating that algal amino acid labeling neither sacrifices sensitivity, nor reduces peak count (Fig. 2.4, A). Proton linewidths are comparable, but slightly broader in ISOGRO® than in ^{2}H/^{15}N/^{13}C (65.3 ± 19.9 Hz vs. 61.4 ± 14.6 Hz), consistent with the higher proton density and faster T_2 decay expected for the more protonated ISOGRO® sample (Fig. 2.4, B). The difference of 4.1 Hz between median proton linewidths is comparable with reported values in solution NMR by Löhr et al. [107]. Thus, ISOGRO® sample maintains the spectral resolution previously attainable only with full perdeuteration in micro-crystalline TS.

To estimate additional resonances introduced by ISOGRO®, we relied on the hCONH experiment [111]. CCPNMR AnalysisAssign 3.2.2 program used here for data processing labels a peak once its intensity declines by ≥10% before hitting a noise floor set to 3.5σ, which nominally corresponds to SNR = 2.8. Applied to the ISOGRO® data, this threshold returned 1225 peaks in ISOGRO® sample, which is twice as many as the number of residues in TS (667 residues from both TrpA and TrpB subunits), indicating extensive picking of noise. To identify a more reliable cutoff, we plotted the cumulative peak count versus SNR and traced where the curve transitions from signal-dominated to noise-dominated growth (Fig. 2.4, D). An empirical (and rather conservative) boundary of SNR = 7.4 reduced the list to 532 peaks, i.e. 98 more than in the perdeuterated control (Fig. 2.4, C). Approximately half of the 196 theoretically missing signals remain invisible, presumably because fast exchange or local mobility broadens them beyond detection. Nevertheless, the 98 additional peaks substantially enlarge the dataset available for backbone assignment.

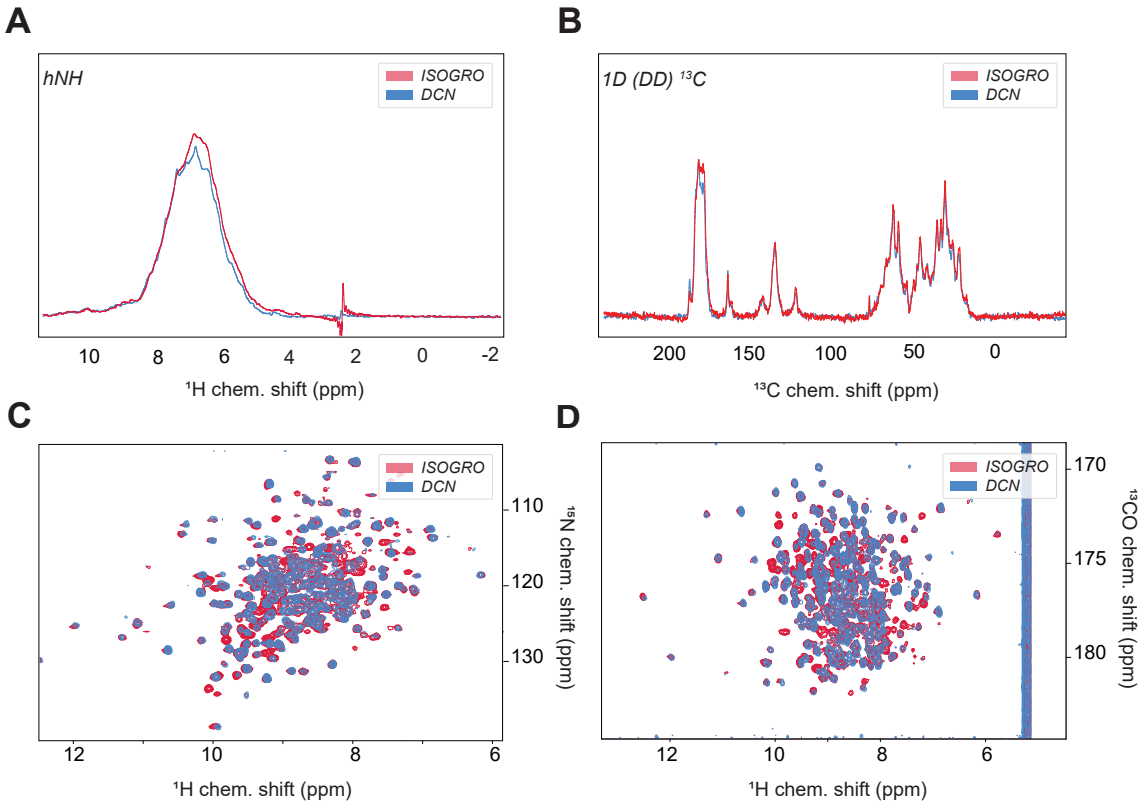


Figure 2.3: Signal intensity and spectra resolution for the perdeuterated/back-exchanged sample ($^{2}H/^{15}N/^{13}C$) and algal amino acids-enhanced sample (ISOGRO®). **(A)** hNH bulk spectra. **(B)** ^{13}C direct polarization spectra used for deriving a scaling factor. **(C)** and **(D)** Overlay of H/N and H/CO correlations for TS obtained at 700 MHz Larmor frequency from a perdeuterated and ^{1}H-back-exchanged sample ($^{2}H/^{15}N/^{13}C$) or algal amino acid sample (ISOGRO®) depicted for an effective proton evolution time of 20 ms and sin^2 apodization with a sine bell shift of $\pi/4$.

2.5.4 Protonation level in ISOGRO® sample

The ISOGRO® hCONH spectrum reveals a heterogeneous pattern: non-polar residues such as Ala, Val and Leu retain <30% total ^{1}H, whereas acidic side chains are substantially reprotonated (Asn, Glu >80%) (Fig. 2.4, E). Average values amount to 26% at $^{1}H_{\alpha}$ and 20% over all other aliphatic positions, corresponding to an overall non-exchangeable protonation of ≈23%. By contrast, the iFD control (Fig. 2.4, F) reaches ≻95% $^{1}H_{\alpha}$ and ≈80% side chain protonation, underscoring the diluted proton network introduced by algal lysate-based labeling.

The elevated proton levels in Asn and Glu likely arise from acid-/base-catalysed hydrolysis during commercial lysate production, which partially replaces deuterons by protons. Employing fully synthetic, uniformly deuterated amino-acid mixes would likely decrease these outliers to the low values observed for Ala, Arg, Lys and Thr.

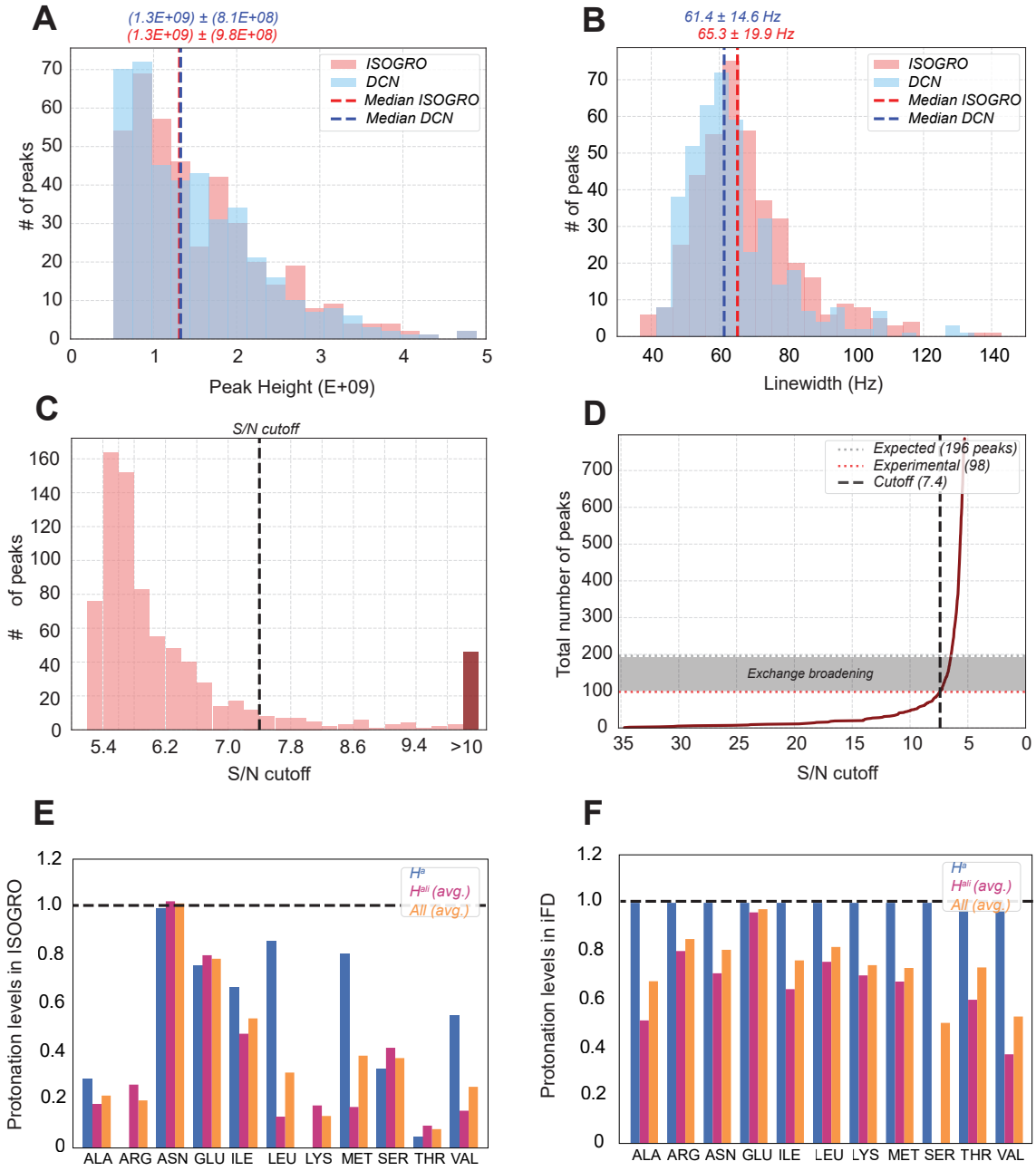


Figure 2.4: Quantitative assessment of the performance of algal amino acid extract as a replacement for traditional glucose- and D_2O-based labeling approach for MAS NMR of large protein complexes. **(A)** Histogram of number of peaks vs. normalized peak height. **(B)** Histogram of number of peaks vs. linewidths (Hz). Vertical markers/labels denote the median ± SD. **(C)** Histogram of number of new peaks emerging in ISOGRO® sample at various SNR values. ≻10 bin comprises all peaks with SNR higher than 10. SNR cutoff value depicted as dashed vertical line was chosen visually and in line with expected number of peaks. **(D)** Cumulative number of new peaks in ISOGRO® sample at various SNR. Expected number of peaks (196) as well as experimentally determined number of peaks (98) are plotted as dotted grey and red lines, respectively. The discrepancy between the expected and observed number of newly emerging peaks is likely due to the chemical exchange broadening in flexible loops of a protein. **(E)** Protonation chart for water/algal extract-based sidechain deuteration (ISOGRO® labeling). **(F)** Quantitative protonation chart for deuteration based on glucose- and H_2O-containing minimal medium (iFD labeling).

Taken together, algal-extract labeling results in a highly protonated backbone while preserving near-perdeuterated side chains. This is an advantageous compromise that maintains narrow MAS linewidths and rescues previously silent amides in the hydrophobic core.

2.6 Discussion

Tested here algal lysate-based selective deuteration revisits an idea first demonstrated by Löhr et al. for solution NMR, in which protein expressed with algal lysate instead of $^{2}H/^{13}C$-glucose and $^{15}NH_4Cl$ displayed a significant increase in number of peaks compared to a standard $^{2}H/^{15}N/^{13}C$-labeling approach [107]. Here we probe algal lysate concept under the following conditions: (a) microcrystalline *Salmonella typhimurium* tryptophan synthase (TS) complex of molecular weight of 144 kDa grown without added glucose or ammonium salt, so the algal amino acids represent the sole carbon- and nitrogen-sources; (b) tight packing of prism-like crystals; (c) data collection at 700 MHz and 55 kHz MAS.

We show that hCONH ^{1}H linewidths center around 65.5 Hz, which is only 4 Hz broader than the perdeuterated reference. Peak heights remain unchanged after correcting for the amount of sample, and every DCN peak is retained. A conservative signal-to-noise sweep (cut-off 7.4) reveals 98 additional resonances, which comprises half of the 196 theoretically missing sites but already sufficient to improve sequential backbone assignment.

Solution-state measurements on the SH3 domain show that ISOGRO® extract yields almost complete protonation of backbone amides and partial protonation of $^{1}H_{\alpha}$ sites while leaving large parts of the side chains strongly deuterated. $^{1}H_{\alpha}$-protonation is not uniform: acidic amino acids are almost fully protonated, that is a pattern that can be traced to acid/base hydrolysis during commercial lysate preparation. The resulting proton distribution differs from the more uniformly labeled iFD samples used as reference. For MAS NMR this mix of high amide proton density and sparse aliphatic protons is attractive, because it boosts sensitivity where it matters most without broadening lines excessively.

In conclusion, deuterated cell lysate media make full amide protonation possible in H_2O, but the practical impact on spectral quality depends on how the sample is grown, packed and spun. In our experiments we used microcrystalline TS sample that possessed a certain degree of crystallization irreproducibility, which we attributed to slightly varying pH and/or temperature upon crystallization and crystals' incubation. Better control over pH and temperature during nucleation, or the use of micro-seeding, should further tighten linewidths and could reveal the remaining missing peaks.

2.7 Limitations and future research

Accurate quantification of the impact of amino acid extracts on TS yield requires a more rigorous and statistically robust experimental workflow. Protein yields should be assessed using independent biological replicates, and should include measurement of protein-to-cell mass ratios. Quantitative analysis should employ more robust methods, such as calibrated protein assays or absolute quantification against

standards. Additionally, the protonation states of amino acids in each extract should be determined quantitatively to facilitate meaningful comparisons and to identify the most effective extract composition. The optimal extract concentration for maximizing H/D exchange without compromising sample quality should be determined empirically by preparing a concentration series and evaluating peak counts, linewidths, and intensities across multiple protein targets.

Chapter 3

TrpB: from crystals to NMR spectra

This chapter is based on work published in [112].

3.1 Introduction

Protein NMR characterization starts with sample preparation. Protein must be purified to homogeneity so that no contaminants or degradation products remain, which can compromise spectral resolution and assignment. For MAS NMR, sample preparation relies either on sedimentation via ultracentrifugation, or on a demanding procedure of protein crystallization. Both methods have their own advantages and established prevalence in the field. Microcrystalline samples typically provide well-ordered, high-density material with spectral features similar to those of diffraction-quality crystals [113][114]. However, crystallization screening are often time-consuming and may not succeed with charged or flexible proteins [115][116] and rigid lattice often renders flexible regions invisible in CP-based experiments [117]. Additionally, rigid lattices can make flexible regions undetectable in CP-based experiments. Sedimented protein samples offer a robust alternative, especially when crystallization is challenging or when the target system involves dynamic, multimeric complexes [118][119]. The limitation of sedimentation in MAS rotors is molecular weight of proteins: the smallest protein to be sedimented is 17 kDa ubiquitin dimer [120].

Compared to microcrystalline MAS samples, sediments frequently show altered relaxation behavior: often shorter ^{1}H T_2 and/or different T_1, depending on nucleus and sample order, consistent with increased fast motions [118][121]. Practically, this can improve spectral resolution but does not automatically translate into shorter recycle delays [119]. Although both sedimentation and microcrystallization are widely used, practical criteria for selecting between them for a specific protein and experiment remain system-dependent and are rarely quantified under comparable conditions.

Sedimentation and microcrystallization represent two established methods for preparing MAS-ready protein samples. Systematic comparison between these approaches requires that each sample preparation is optimized and remains stable under the conditions necessary for multidimensional MAS NMR. In practice, microcrystalline samples are frequently used in combination with paramagnetic doping to

shorten T_1 time and hence acquisition time [122]. However, a persistent issue with doping of protein crystals exists: introducing metal complexes can perturb crystal integrity, for example by altering ionic strength or by chelating lattice-stabilizing ions [123][124][125]. Several approaches were tested to tackle to issue of dissolving crystals, such as alternative chelating cages for Cu^{2+} ion, gradual introduction of a dopant into the mother liquor, co-crystallization of a target protein in the buffer containing dopant, and PEG coating prior to doping, with the latter being the most successful method that resulted in crystals stabilization for up to four weeks [123]. Since the effectiveness of these methods is often sample- and condition-dependent, additional stabilization strategies need to be developed. One method of interest that has not yet been tested for MAS NMR is chemical cross-linking of crystals. It can be combined with controlled dopant exposure to improve microcrystal robustness while maintaining MAS NMR spectral quality.

3.2 Aim of the project

This chapter aims to establish a practical workflow for preparing TrpB samples suitable for multidimensional MAS NMR and for generating backbone resonance assignments from combined solution- and MAS NMR datasets. Specifically, we (a) compared batch and vapor-diffusion crystallization routes with respect to spectral quality, (b) evaluated crystal stabilization by glutaraldehyde cross-linking under paramagnetic doping conditions, (c) benchmarked sedimented, microcrystalline, and solution samples using quantitative spectral metrics, and (d) used complementary solution- and MAS NMR experiments to obtain partial backbone assignments and to assess dataset reliability using overlap, connectivity-ambiguity, and solution–MAS chemical shift deviation analyses.

3.3 Case study: TrpB from *Pyrococcus furiosus*

Thermophilic proteins exhibit significant benefits over their mesophilic orthologs when used as model objects in structural biology. One of the primary advantages is their inherent enhanced thermostability, which leads to improved solubility, crystallization propensity, and overall stability during experiments such as X-ray crystallography and NMR spectroscopy [126]. These attributes allow for more reliable and high-resolution structural data acquisition. For instance, proteins from thermophilic organisms are often more soluble and produce crystals with superior diffraction properties, featuring lower B-factors and improved crystallographic R-factors [127]. These advantages motivated the choice of *Pyrococcus furiosus* TrpB (PfTrpB) for the present study, as its parent enzyme shows superior kinetic parameters and thermal robustness compared with other bacterial or archaeal homologues.

Tryptophan synthase is an $\alpha\beta\beta\alpha$ heterotetramer that channels indole over a 25 Å hydrophobic tunnel from the α-subunit (TrpA), which cleaves indole-3-glycerol phosphate, to the β-subunit (TrpB), which joins indole with L-serine to form L-tryptophan [128]. The two reactions are kept in phase by allosteric interactions between the two subunits, mediated by the discrete networks of residues that are involved in signal propagation upon effector binding. Within TrpB, the PLP cofactor cycles through a sequence of covalent intermediates. After external aldimine forma-

tion E(A_{ex1}), deprotonation gives the quinonoid E(Q_1) followed by elimination of water to yield the amino-acrylate E($A-A$). Reaction of indole with E($A-A$) forms a second quinonoid that is finally protonated to release product. This eight-step sequence is gated by a rigid-body motion of the COMM domain (communication domain), which swings from open to closed conformations as chemistry proceeds (Fig. 3.1).

In the absence of TrpA, TrpB energy landscape is dominated by open or partially closed states that accumulate early intermediates like E(A_{ex1}) and limit turnover up to 5% of the activated complex [129]. Laboratory evolution introduced eight remote mutations (PfTrpB^{2B9}) that restore and even exceed the catalytic efficiency of the $\alpha\beta\beta\alpha$ complex (Fig. 3.1, A, B) [130]. Structural and spectroscopic analyses show that these mutations increase the population of closed conformers, increase flexibility in the COMM domain and accelerate progression to E($A-A$) and quinonoid states without perturbing catalytic residues. The engineered enzyme, therefore, emulates the effect of TrpA binding, while remaining a monomeric, substitution-tolerant scaffold.

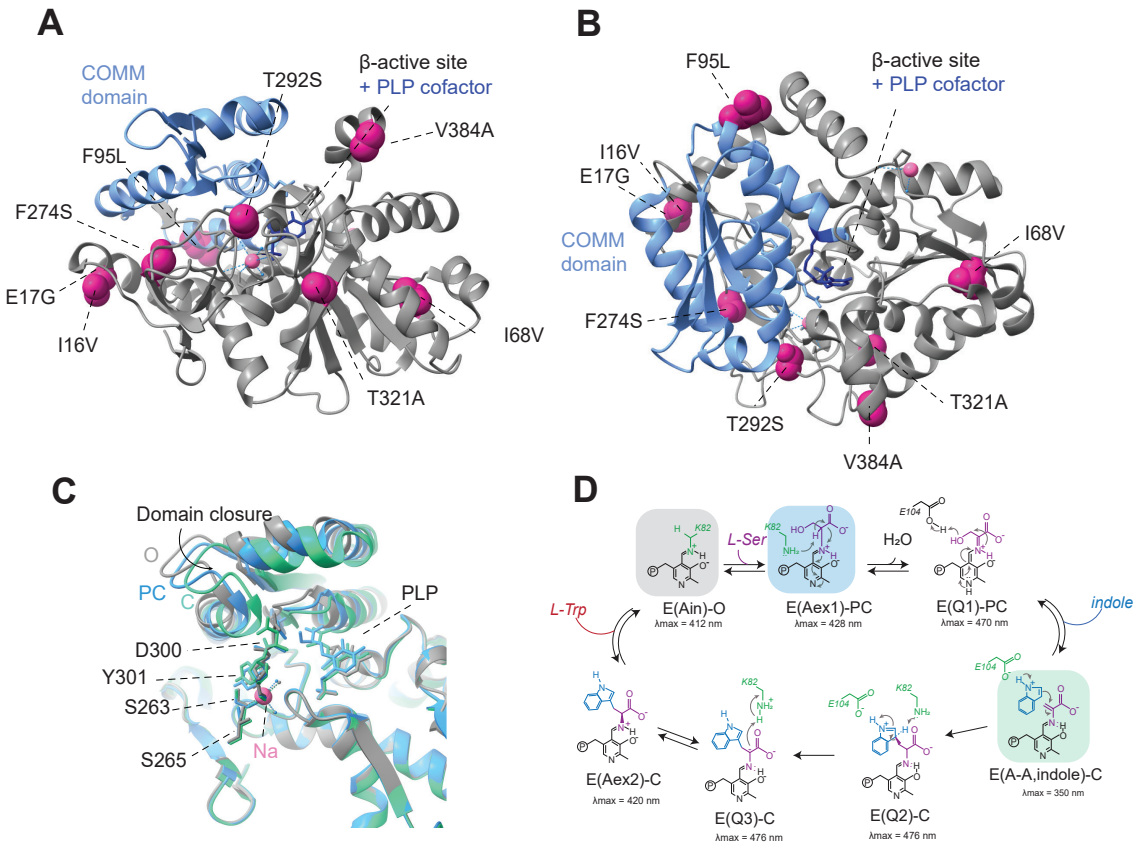


Figure 3.1: Structural organization of PfTrpB. **(A)** and **(B)**. Distribution of activating mutations in a standalone variant of TrpB named PfTrpB^{2B9} identified through random mutagenesis and screening by A. Buller [131] The highlighted mutations are I16V, E17G, I68V, F95L, F274S, T292S, T321A, and V384A. **(C)** Structural transitions around and in COMM domain upon L-Ser binding in PfTrpB. Superimposed are three structures: the open state in the absence of a ligand is in gray (PDB: 5DVZ). The external aldimine of L-Ser, E(A_{ex1}), is observed with a partially closed conformation in blue (PDB: 6AMH). The formation of the electrophilic E($A-A$) intermediate is coupled to the formation of a fully closed conformational state in green (PDB: 5VM5). **(D)**. The mechanism of TrpB catalysis. Transimination of L-Ser to form an external aldimine, E(A_{ex1}), followed by deprotonation gives the quinonoid intermediate E(Q_1), is designated stage I of the overall reaction. It is followed by elimination of water to yield the amino-acrylate E($A-A$). Reaction of indole with E($A-A$) forms a second quinonoid E(Q_2) that is eventually protonated to release the product. This eight-step sequence is gated by a rigid-body motion of the COMM domain, which swings from open to closed conformations as chemistry proceeds.

PfTrpB is naturally thermostable and expresses well in *E. coli*, giving multi-milligram yields of soluble protein. Together, these observations place PfTrpB at the intersection of two themes central to this chapter: (a) preparation of high-quality NMR samples and wet-lab method development, and (b) backbone resonance assignments using the combination of solution- and MAS NMR samples.

3.4 Materials and methods

3.4.1 Protein expression and purification

Details on buffer composition can be found in Supplementary Table A.4.

For all labeled expressions, modified M9 media was used that has double amount of buffer salts. The gene encoding PfTrpB (UNIPROT ID: Q8U093) was previously codon-optimized for *E. coli* BL21 (DE3) and cloned into pET22(b)+ with an uncleavable C-terminal 6xHis tag [131]. Expression and purification protocols for the variants reported here were similar to those reported previously. Briefly, a single colony of *E.coli* BL21 (DE3) cells harboring the PfTrpB plasmid was used to inoculate a 5-mL culture of LB media with 100 μg/mL carbenicillin (LB_{carb}) and incubated overnight at 37°C at 180 rpm. This culture was used to inoculate a 50 ml LB_{carb}, which was incubated at 180 rpm and 37°C for 4 hours. Then, 500 μl of culture was inoculated in 50 ml of modified M9 media dissolved in water ($M9_{mod}/H_2O$), supplemented with carbenicillin, and incubated at 180 rpm and 37 °C for 4 hours. After that, 500 μl of culture was inoculated in 50 ml of a composite minimal media mixture comprised of 10% $M9_{mod}/D_2O$ and 90% $M9_{mod}/H_2O$, supplemented with antibiotic, and incubated at 180 rpm and 37°C overnight. Next morning, 500 μl of culture was inoculated in 50 ml of media 50% $M9_{mod}/D_2O$ and 50% $M9_{mod}/H_2O$ and incubated at 180 rpm and 37 °C for 4 hours. Finally, 500 μl of a resultant culture was inoculated in 5 L flask containing 2 L of $M9_{mod}/D_2O$ media supplemented with carbenicillin, and incubated at 180 rpm and 37 °C until OD_{600} reached 1. Cultures were chilled on ice for 20-30 min and expression was induced by the addition of IPTG to a final concentration of 1 mM. Cells continued to grow in a shaker-incubator at 180 rpm and 20 °C for another 20 h. Cells were harvested by centrifugation at 4 °C and 4,000g for 20 min; the pellets were frozen at -80°C until further use.

Frozen cell pellets were thawed at room temperature and resuspended in **buffer L** supplemented with 200 μM PLP, 1 mM PMSF, 1 mM TCEP, 1 mg/mL hen egg white lysozyme, and 0.02 mg/mL DNAse I. After vortexing, cells were lysed with Emulsiflex C3 (three runs for each culture). Lysates were aliquoted into 45 ml tubes and centrifuged at 100,000 g and 4°C for 20 min, then incubated at 75°C for 10 min and centrifuged again as described above. The cleared, heat-treated lysate was applied to a 5 ml Ni-NTA column via gravity filtration. $PfTrpB^{wt}$ and $PfTrpB^{2B9}$ were eluted with **buffer E**. Purified enzyme was exchanged into **buffer N** using PD10 column, frozen in liquid nitrogen, and stored at -80°C until further use.

For solution NMR measurements, PfTrpB protein was centrifuged at 20,000 g for 5 min at 4°C, concentrated to 0.8-1 mM, and mixed with 10% (V/V) D_2O and 50 μM DSS. For sedimentation in 0.7 mm rotor, PfTrpB protein was produced in $^{1}H/^{15}N/^{13}C$-labeling fashion (using perdeuterated and ^{13}C-labeled glucose, ^{15}N-labeled NH_4Cl, and H_2O as a solvent) and sedimented for 46 h in a Beckmann

SW 40 Ti rotor at 20,000 rpm (71,100 g) and 20°C. For sedimentation into 1.3 mm rotor, protein was produced in triple-labeled fashion as $^{2}H/^{15}N/^{13}C$ protein (perdeuterated and ^{13}C-labeled glucose, ^{15}N-labeled NH_4Cl, and D_2O as a solvent), and sedimented for 44 h in a Beckmann SW 40 Ti rotor at 20,000 rpm (71,100 g) and 20°C. For MAS NMR, $PfTrpB^{wt}$ protein was exchanged into **H-buffer** and concentrated to 8-12 mg/ml. Protein was crystallized by vapor diffusion in sitting drop fashion by mixing 1:1 protein solution with **C2 buffer**. Crystals grow after 1 week of incubation at 20°C. For batch crystallization, $PfTrpB^{wt}$ was concentrated to 10 mg/ml in 50 mM potassium phosphate buffer buffer with pH = 8.0, and mixed with the ratio 1:1 with **C3-buffer** in the Eppendorf tube to a final volume of 400 μl. Crystals formed immediately.

3.4.2 Crystals cross-linking

Crystals cross-linking was tested as (a) a gentle vapor diffusion (200 μl 25% Glutaraldehyde (GA) in the crystallization plate mother liquor well for 30 min), and (b) harsh cross-linking via mixing 1:1 (V:V) protein crystals in mother liquor with 25% GA, incubation for 1 min followed by crystals wash with mother liquor. To test crystals stability upon cross-linking, they were mixed with 5 μl of 350 mM Cu-EDTA pH=6.47 to a final concentration of Cu-EDTA c=175 μM. Crystals morphology was assessed via light microscope after 30 min, 60 min, and 3 days of incubation with Cu-EDTA.

3.4.3 NMR data acquisition

For the comparison of microcrystalline, sedimented, and solution PfTrpB samples, perdeuterated and uniformly $^{15}N/^{13}C$-labeled PfTrpB was obtained in three physical states: microcrystalline, centrifugal sediment, and dilute solution. All solid-state samples were packed in corresponding MAS rotors and hNH and hCONH spectra were recorded at 55.5 kHz MAS and 700 MHz ^{1}H frequency with the number of scans for microcrystalline sample NS = 8, and for sedimented sample NS=16. The solution reference was recorded at 800 MHz with a TROSY-HMQC module with number of scans NS=8. Peak amplitudes in hCONH spectrum recorded for a sedimented sample were divided by 2 to account for the double number of scans. Matched data sets were retained for quantitative analysis: a 2D hNH (solid samples) or HMQC-TROSY (solution sample) series for visual assessment of peak count and dispersion, and a 3D hCONH / HNCO series for numerical linewidth and peak-height statistics. Processing parameters for MAS samples were kept identical within each dimension (EM with LB=50 Hz in ^{1}H dimension, QSINE with SSB=4 in ^{15}N and ^{13}C dimensions).

For solution NMR $PfTrpB^{2B9}$ assignment all NMR spectra were recorded at 45°C on a Bruker Avance NEO 800 MHz ^{1}H Larmor frequency spectrometer equipped with a CP TCI proton-optimized triple-resonance cryoprobe. The recorded experimental dataset consists of the following experiments: 2D TROSY-HSQC, 3D TROSY-HNCA, 3D TROSY-HNCO, 3D TROSY-HNcaCO, 3D TROSY-HNcoCA, 3D HN-CACB. For MAS NMR assignment experiments of crystalline $PfTrpB^{wt}$ 3D experiments were recorded at 55.5 kHz MAS and 700 MHz ^{1}H frequency. Experiments 4D shCACONH and 4D shCOCANH were recorded as described earlier [85] with

TROP-pulses for sensitivity enhancements as described in [59][54], and 4D hCACB-caNH was recorded in a standard fashion. The full list of experiments recorded for this chapter can be found in Supplementary Table A.6. Sensitivity-enhanced 4D MAS experiments were recorded in NUS fashion with 1.04373% points of sparse sampling and 4000 of hypercomplex points in indirect dimension. 4D hCACBcaNH experiment was recorded in NUS fashion with 2.07679% points of sparse sampling and 10532 of hypercomplex points in indirect dimension. NUS reconstruction was performed with SSA algorithm [132] implemented within the in-home developed python script.

3.4.4 Estimation of water content from 1D ^{1}H spectra

The relative water content in microcrystalline and sedimented PfTrpB MAS samples was estimated from integrated 1D ^{1}H MAS NMR spectra (Fig. 3.4, A). Spectra were baseline-corrected and integrated in three regions: the water resonance (4.7-5.8 ppm) and two protein-dominated regions on either side of the water peak (upfield and downfield). The sum of the two protein-region integrals was taken as the total protein ^{1}H signal, I_{prot}, while the 4.7-5.8 ppm integral was taken as the water signal, I_{water}. To convert integral ratios into a water/protein mass ratio, the protein signal was referenced to the expected number of non-exchangeable protons per PfTrpB molecule, $N_{\mathrm{H,prot}} = 4862$. Assuming two protons per water molecule, the water-to-protein mass ratio (w/w) was calculated as

$$\frac{m_{\mathrm{water}}}{m_{\mathrm{protein}}} = \left(\frac{I_{\mathrm{water}}/2}{I_{\mathrm{prot}}/N_{\mathrm{H,prot}}}\right)\left(\frac{M_{\mathrm{water}}}{M_{\mathrm{protein}}}\right), \tag{3.1}$$

, where $M_{\mathrm{water}} = 18$ g mol^{-1} and $M_{\mathrm{protein}} = 66843$ g mol^{-1}. Water-to-protein ratios were determined for freshly prepared samples and after storage in the spectrometer for the indicated durations.

3.4.5 Data processing, analysis, and plotting

NMR data were processed using TopSpin 4.5.0 (Bruker) and analyzed using CCPNMR Analysis v. 3.2.4 (https://ccpn.ac.uk/)[110]. For processing and plotting, custom in-home Python scripts were used.

To quantitatively compare hCONH spectra recorded for PfTrpB samples crystallized in sitting-drop- and batch-manner, only peaks present in both samples were considered significant. A manual threshold was chosen to distinguish between peaks and noise. Linewidth distributions were analyzed using Kernel Density Estimate (KDE) to generate probability density functions. The distribution of linewidths in sedimented, crystalline, and solution sample preparations were compared across all three dimensions using 1D KDE plots with curves normalized to equal area. Peak height distributions were analysed using KDE plot with curves normalized to equal area. Peak heights from the sedimented sample were scaled down by a factor of 2 to account for the doubled number of scans used during acquisition, ensuring comparable signal intensities across datasets. Peak displacement between sedimented and solution, and microcrystalline and solution samples was calculated using residue-matched peaks present in all three samples. Chemical shift displacements were calculated as the scalar sum of absolute differences in each dimension relative to the

solution reference: $\Delta = |\Delta\delta^{1}\mathrm{H}| + |\Delta\delta^{15}\mathrm{N}| + |\Delta\delta^{13}\mathrm{C}|$, where $\Delta\delta$ represents the chemical shift difference (MAS minus solution) for each nucleus type. Height-weighted kernel density estimation was performed as usual. To quantify peak overlap between sample preparations, a three-way Venn diagram was constructed with matplotlib-wenn package.

For backbone resonance assignment, solution-derived chemical shifts enriched with MAS-derived chemical shifts along with chemical shift predictions from ShiftX2 [133] and an X-ray structure of a ligand-free $\mathrm{PfTrpB}^{2\mathrm{B9}}$ (PDB ID: 6AM7) were used for automated assignment in FLYA [134] followed by manual inspection. Assignments were classified as strongly supported when they were reproduced in more than 80% of independent algorithm runs. A list of partial assignments can be found in Supplementary Table A.7.

3.5 Results

3.5.1 Crystallization in batch and in slow vapor diffusion manner

We attempted to crystallize both $PfTrpB^{wt}$ and $PfTrpB^{2B9}$. Crystallization conditions for both are known (8-12 mg/ml $PfTrpB^{wt}$ in 20 mM HEPES pH=7.85 mixed with **crystallization buffer C2** (100 mM Na-HEPES pH=7.85, 15-20% PEG3350), with the only difference being that $PfTrpB^{2B9}$ crystallization requires cross-seeding with seeds of $PfTrpB^{wt}$ [130]. In our laboratory setup, $PfTrpB^{2B9}$ did not crystallize likely due to the inherent flexibility, and the standalone mutant was therefore later studied in solution. $PfTrpB^{wt}$ was routinely crystallized both in batch manner (mixing 1:1 V/V protein with **crystallization buffer C3** in large volumes in the Eppendorf tube followed by immediate formation of homogenous microcrystals) and in slow vapor diffusion manner (sitting drop and hanging drop) (Fig.3.2, D).

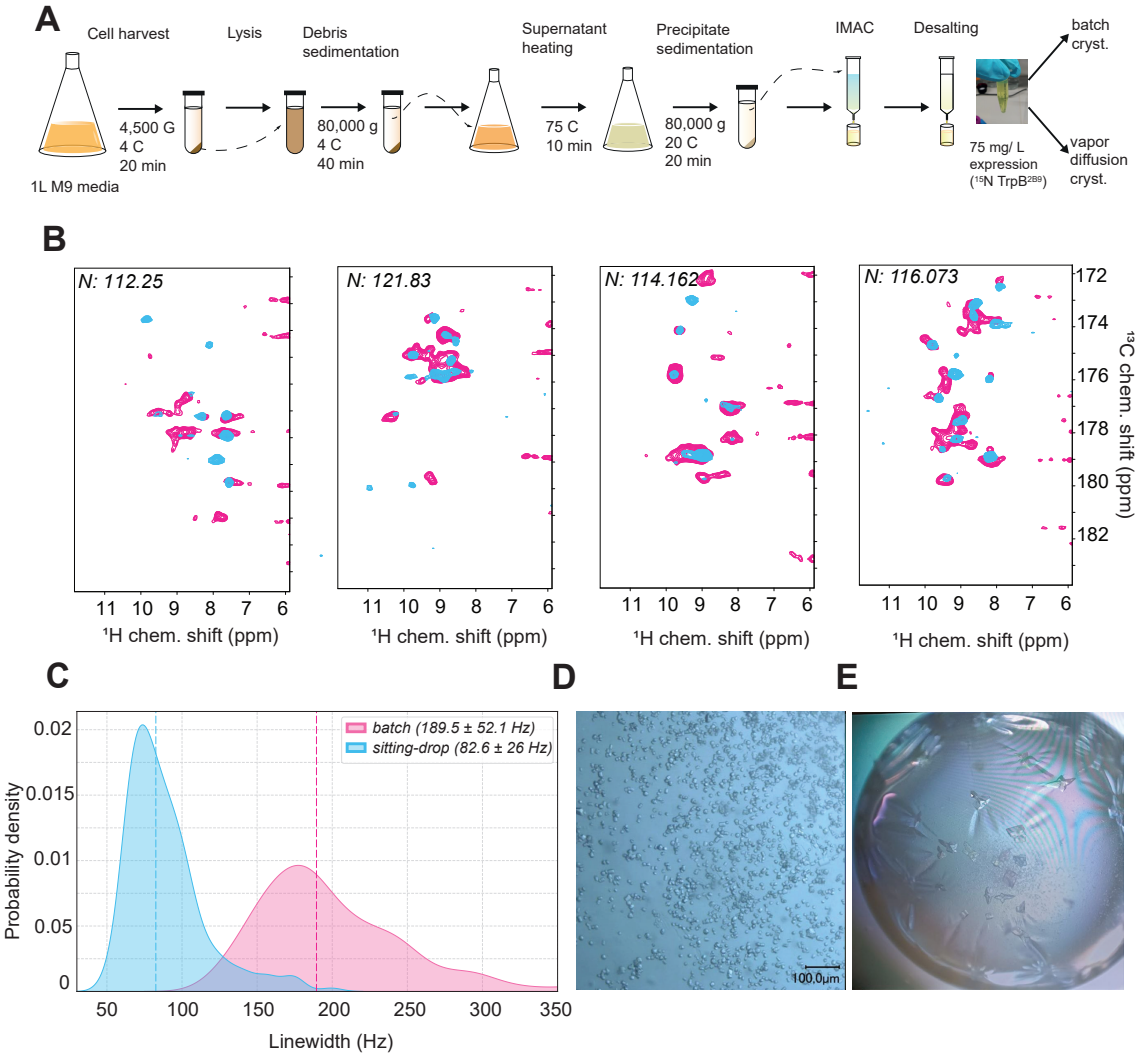


Figure 3.2: PfuTrpB sample preparation and testing. **(A)** Graphical representation of the purification of PfTrpB using gravity filtration Ni-NTA column followed by buffer exchange with PD10 column. Natural hyperthermostability of *Pyrococcus* proteins allows for the highly efficient heat precipitation of all contaminant proteins from *E.coli* BL21 (DE3). **(B)** 2D planes at different ^{15}N coordinates obtained from hCONH spectra recorded on batch crystallized PfTrpB^{wt} in buffer C3 (pink) and sitting-drop crystallized PfTrpB^{wt} in buffer C2 (blue). Broader linewidths and more noise in batch crystallized sample show that batch crystallization for PfTrpB^{wt} does not provide backbone assignment-quality spectra. **(C)** KDE plot of ^{1}H linewidths of batch-crystallized PfTrpB^{wt} (pink) and sitting-drop-crystallized PfTrpB^{wt} (blue). Median values are 189.5 Hz and 82.6 Hz, respectively. **(D)** Microcrystals of batch-crystallized PfTrpB^{wt}. **(E)** Microcrystals of sitting-drop-crystallized PfTrpB^{wt}.

For MAS NMR purposes, batch protein crystallization poses a big advantage of a relatively fast and easy sample preparation and is a more common way to prepare microcrystalline samples than slow vapor diffusion in crystallization plates [42][85][86]. However, in case of $PfTrpB^{wt}$ batch-crystallized protein resulted in severely broadened linewidths compared to the slowly growing in vapor diffusion plates crystalline protein (Fig. 3.2 B, C) – 189.5 Hz and 82.6 Hz, respectively. We suggest that fast crystallization in batch manner results in heterogenous mixture of crystals, which, in turn, causes severe line-broadening. We therefore continued working with slowly growing crystals.

3.5.2 Cross-link series

To address the dissolution of crystals during paramagnetic doping, glutaraldehyde (GA) was evaluated as a cross-linking agent. The initial sample was prepared using a stringent protocol by mixing crystals and 25% GA in a 1:1 (V/V) ratio for 1 min, followed by washing with mother liquor. After treatment, the crystals demonstrated increased adhesion to plastic surfaces, which complicated handling. The cross-linked crystals remained intact in the presence of Cu-EDTA for at least 3 days. However, the 2D hNH spectrum exhibited pronounced line-broadening, with only a limited number of peaks detectable (Fig. 3.3, C), indicating a significant loss of spectral resolution.

We therefore tested a milder protocol in which the batch solution was replaced with 25% GA and the plate was incubated for 30 minutes to allow GA vapor to diffuse into the crystals within the well (Fig. 3.3, A) [135]. Crystals treated using this approach demonstrated comparable long-term stability and resisted dissolution during a 3-day Cu–EDTA exposure (Fig. 3.3, B), as well as resuspension in pure deionized water and concentrated 37% HCl acid. Nevertheless,the corresponding hNH spectrum, while somewhat improved compared to the initial protocol, still exhibited severely broadened linewidths in both dimensions.

Taken together, these findings indicate that GA cross-linking under the tested conditions stabilizes crystals against dopant-induced dissolution but reduces spectral quality, likely as a result of chemical heterogeneity or structural perturbations caused by extensive cross-link formation. Subsequent optimization efforts should prioritize the evaluation of alternative cross-linking agents and the implementation of more controlled reaction conditions.

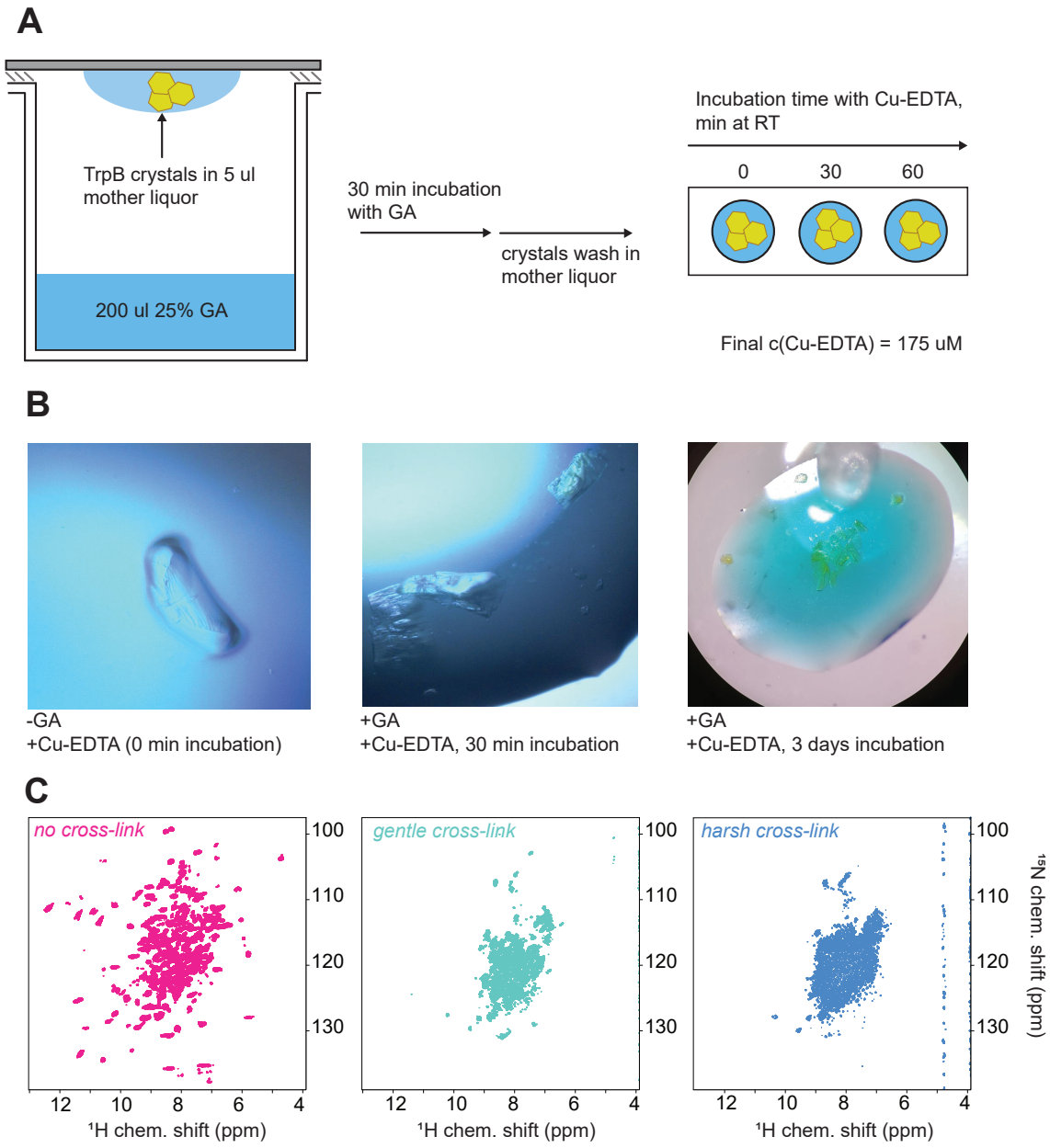


Figure 3.3: Protocol for and results of GA-mediated PfTrpB^{wt} protein crystals cross-linking. **(A)** Experimental design of a 'gentle' vapor diffusion cross-linking followed by the crystals wash and mixing with Cu-EDTA. Presence or absence of dissolving crystals was assessed with the microscope. A typical sign of dissolution is blurring of crystals' edges right after addition of Cu-EDTA, with no crystals present in the well after 5 min of incubation. **(B)** Crystal morphology without GA (control sample, left), and with GA after 30 min of incubation with Cu-EDTA (center) and after 3 days of incubation with Cu-EDTA (right). **(C)** 2D hNH spectra recorded on a control sample with no GA (left), with the gentle vapor diffusion cross-linking (center) and Cu-EDTA doping, and with harsh cross-linking and Cu-EDTA doping (right).

Although stabilized protein crystals were not obtained in this study, the potential applications of such samples are extensive. Once developed, these crystals would enable the investigation of protein interactions with organic solvents, acids, and bases, and facilitate the introduction of beneficial yet frequently destabilizing compounds such as Cu-EDTA.

3.5.3 Sediment vs. microcrystals

The strengths and limitations of various MAS sample-preparation methods were evaluated by comparing solution, sedimented, and microcrystalline preparations of perdeuterated PfTrpB using a standardized set of 2D- and 3D-experiments. The 2D hNH /HMQC-TROSY spectrafacilitated visual assessment of resolution and chemical shift dispersion among samples, with the solution spectrum providing a reference for chemical shift consistency. The 3D hCONH/HNCO data sets enabled a quantitative comparison of peak counts, linewidths in all three dimensions, and the distribution of peak amplitudes across the different samples.

Microcrystalline and sedimented PfTrpB samples displayed distinct time-dependent changes in water content. The microcrystalline sample showed a water-to-protein mass ratio of 9.34 (w/w) at the beginning of the measurements, which decreased to 9.05 after 10 days in the spectrometer. In contrast, the sedimented sample began at 9.85 (w/w) and declined markedly to 6.45 after 8 days. Therefore, the sedimented preparation displays a substantially greater change in hydration over time. This variability in the water content may compromise long-term sample stability of the sample and may contribute to the reduced peak detectability in the 3D hCONH spectrum.

A comparison of the 1D ^{1}H spectra with water suppression demonstrated that the sedimented PfTrpB sample exhibits 6.6-fold higher signal intensity than the microcrystalline PfTrpB (Fig. 3.4, A). To determine whether this difference results from polarization transfer efficiency, the first FID of the hCONH experimnet was examined under identical conditions, revealing comparable transfer efficiency between the two preparations (Fig. 3.4, B). Bulk T_1 relaxation times differed by nearly two-fold, with faster relaxation observed in the microcrystalline sample (Fig. 3.4, C). These findings indicate that the higher 1D signal intensity of the sedimented sample cannot be attributed to more favorable T_1 recovery. Therefore, a higher effective amount of protein likely contributes to the detected signal in the sedimented sample, potentially due to differences in filling factor. However, the longer T_1 in the sedimented sample imposes an additional time penalty for extended experiments.

The sedimented sample yielded a well-resolved ^{1}H-^{15}N correlation spectrum with a CP-based hNH pulse sequence. Relative to the microcrystalline sample, the sediment shows higher degree of line broadening (Fig. 3.4, D). Comparison of peak counts between three samples from hCONH and HNCO spectra showed 286 peaks for sedimented, 326 peaks for microcrystalline, and 309 peaks for solution PfTrpB samples. Since the S/N threshold for peak detection was chosen manually, it is possible that the factual number of peaks in samples is different.

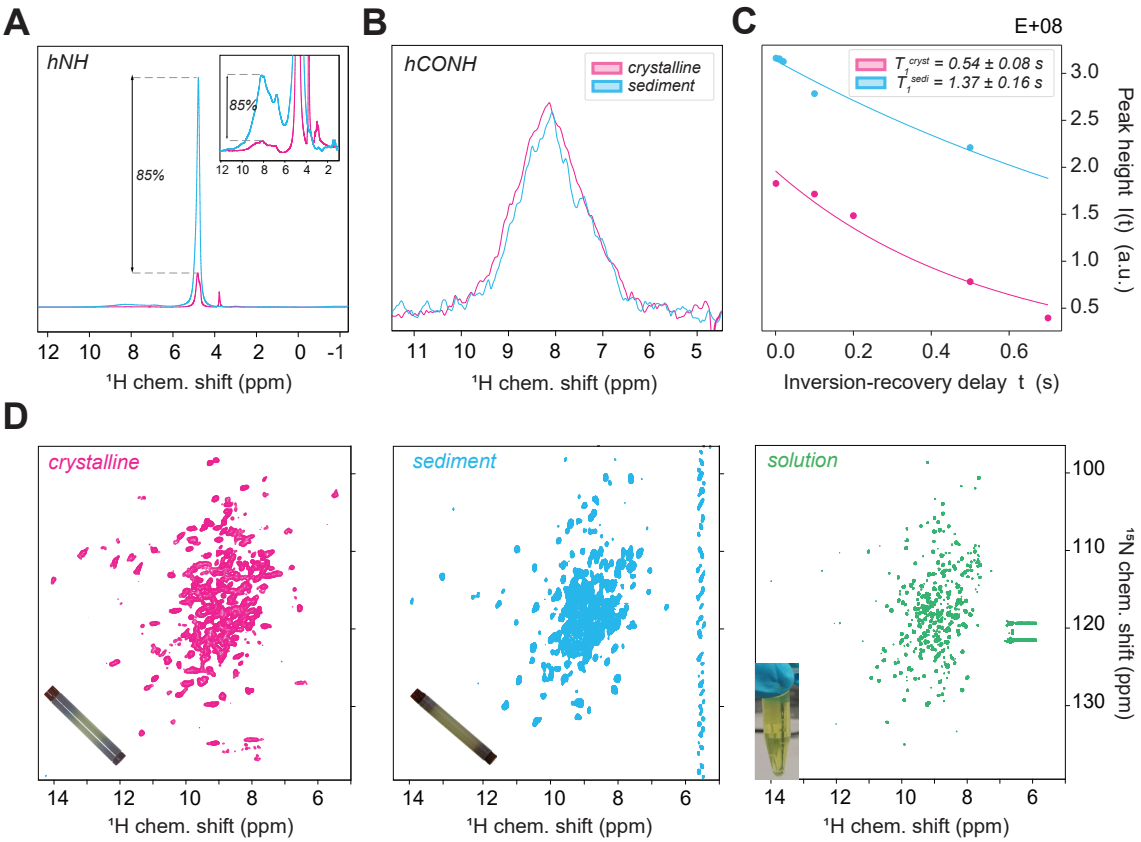


Figure 3.4: Comparison of sedimented, crystalline, and solution spectra of PfTrpB. **(A)** Bulk ^{1}H signal of sedimented and crystalline PfTrpB^{wt} samples, insert spectrum - scaled up region containing protein signals. **(B)** First FID of hCONH experiment for sedimented and crystalline PfTrpB^{wt} scaled to the number of scans. **(C)** bulk T_1 for sedimented and crystalline PfTrpB^{wt}. **(D)** ^{1}H-^{15}N correlation of ^{2}H/^{15}N/^{13}C-labeled PfTrpB in crystalline (pink) and sedimented (blue) versions. Solution ^{1}H-^{15}N correlation recorded as TROSY-HMQC (green) is shown as a reference.

The KDE plots of linewidths show that ^{1}H linewidths cluster at around 34 Hz in solution, and shift to a median of 90 Hz in the sediment and 83 Hz in the crystal, with the latter two exhibiting pronounced tail toward larger linewidths reaching ≈150 Hz (Fig. 3.5, A). Nitrogen ^{15}N linewidths remain narrowest in solution, increase modestly in the sediment, and further expand in the crystalline sample (32, 43, and 56 Hz, respectively) (Fig. 3.5, B). This difference may reflect a less rigid environment in the sediment, for example increased local motions and/or reduced static disorder compared to microcrystals [119]. Carbonyl ^{13}CO linewidths show the smallest inter-sample spread, rising from 47 Hz (solution) to 50 Hz (sediment) and 59 Hz (crystalline), suggesting that the dominant broadening contributions for ^{13}CO are less sensitive to the change in sample state under these conditions (Fig. 3.5, C).

Peak height KDEs share a common intensity mode at 9×10^{8} a.u. with the crystalline distribution decaying fastest, whereas the solution spectrum extends to the highest amplitudes (Fig. 3.5, D). Using the solution coordinates as reference, the median height-weighted displacement $\Delta = |\Delta^{1}\mathrm{H}| + |\Delta^{15}\mathrm{N}| + |\Delta^{13}\mathrm{C}'|$ is 0.23 ppm for the sediment and 0.39 ppm for the crystal, underscoring the more solution-like chemical environment in the sediment (Fig. 3.5, E). Finally, a scaled Venn diagram shows that 184 peaks are common to all three preparations (48% of the total number of expected peaks), while a further 224, 210 and 209 peaks are shared pair-wise between solution–sediment, solution–crystal and sediment–crystal, respectively; 59, 37 and 40 peaks are unique to the individual spectra (Fig. 3.5, F). These unique peaks may reflect genuine chemical shift perturbations or differences in peak detectability near the threshold. These factors complicate direct peak-to-peak matching across samples.

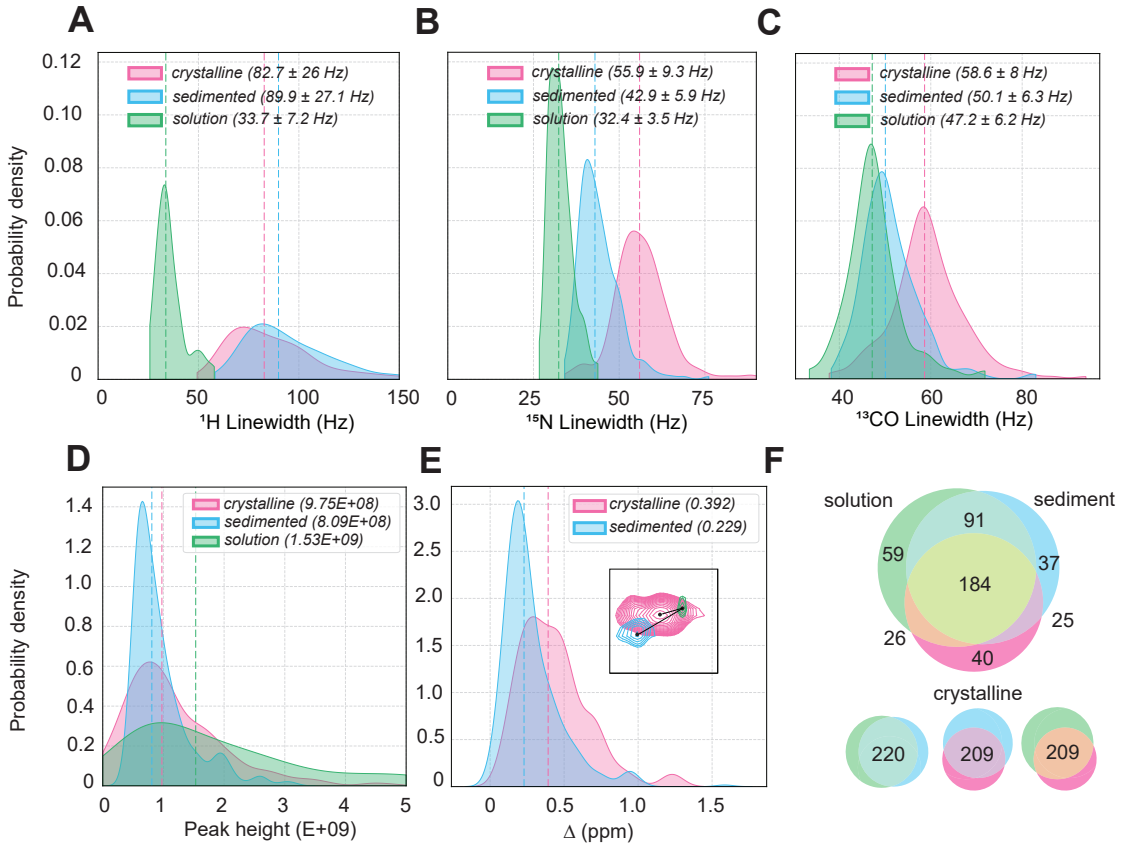


Figure 3.5: Statistical comparison of sedimented, crystalline, and solution spectra of PfTrpB. **(A, B, C)** Kernel density estimates of ^{1}H, ^{15}N, and ^{13}C linewidths in crystalline, sedimented, and solution NMR samples. **(D)** 1D KDE of peak height distribution in crystalline, sedimented, and solution NMR samples. **(E)** Height-weighted 1D KDE of the chemical-shift displacement $\Delta = |\Delta\delta^{1}\mathrm{H}| + |\Delta\delta^{15}\mathrm{N}| + |\Delta\delta^{13}\mathrm{C}|$ of the two solid preparations relative to the solution reference. **(F)** Scaled Venn diagram of peak-set membership. 184 peaks in hCONH spectrum are common for all three samples, 220 peaks are shared by solution and sedimented samples, 209 peaks are shared between sedimented and crystalline samples, 209 peaks are shared between solution and crystalline samples.

Overall, sedimented samples offer a more solution-like chemical environment and higher bulk signal intensity; however, they exhibit greater time-dependent hydration changes and broader ^{1}H linewidths, which may restrict the duration of multidimensional acquisitions. In contrast, microcrystals yield higher peak counts and enhanced hydration stability, although they introduce larger chemical shift perturbations relative to sediments.

3.5.4 NMR assignments: combining solution- and solid-state NMR

To gain access to site-specific information about the protein, the first (and rate-limiting) step is to assign each resonance to an amino acid residue. This chapter combines solution- and MAS NMR approaches to obtain backbone assignments of PfTrpB. For solution NMR assignments of PfTrpB^{2B9}, the high thermal stability of the protein was leveraged by recording all assignment experiments at 45°C to improve spectral resolution. For MAS NMR assignments, PfTrpB^{wt} crystals were used, as the wild-type protein crystallizes more easily and reproducibly than the PfTrpB^{2B9} mutant. Because even minor mutations and changes in sample state can cause chemical shift perturbations, solution assignments served as an anchor, and MAS correlations were accepted only when supported by consistent H/N coordinates and the expected ^{13}CO/^{13}C$_{\alpha}$ connectivities across multiple experiments.

Backbone nuclei resonances (i.e., ^{1}H, ^{15}N, ^{13}CO, ^{13}C$_{\alpha}$, ^{13}C$_{\beta}$) were partially assigned using the 2D- and 3D triple-resonance experiments in solution, as well as 3D- and 4D-experiments in solid-state. The initial round of automated assignments with this approach yielded strongly supported assignments of only 17 residues out of 369 non-proline residues expected. Incorporating these initial assignments with manual backbone assignments performed from both ends of assigned stretches as an additional input, backbone assignments of 147/369 residues (39.8% of all non-proline residues) were obtained. The final round of assignments utilized solution NMR spectra HNCO (314 peaks visible out of 380 expected, or 314/380, which comprises 82.6%), HNCA (511/761, 67.1%), HNcoCA (336/380, 88.4%), HNcaCO (329/761, 43.2%), and HNCACB (176/1437, 12.2%) (Supplementary Fig. B.2). For experiments with low sensitivity and low peak count (3D HNcaCO and 3D HNCACB), analogous 4D MAS NMR spectra (shCOCANH and hCACBcaNH, respectively) were used to enrich the spin system library and facilitate backbone walk. MAS-only assignments based on 3D hCONH, 3D hCANH, 4D shCOCANH, 4D shCACONH, and 4D hCACBcaNH experiments remained limited due to peak overlap and ambiguous sequential connectivities. Incorporation of solution-derived H/N resonances substantially reduced ambiguity and enabled longer backbone walks.

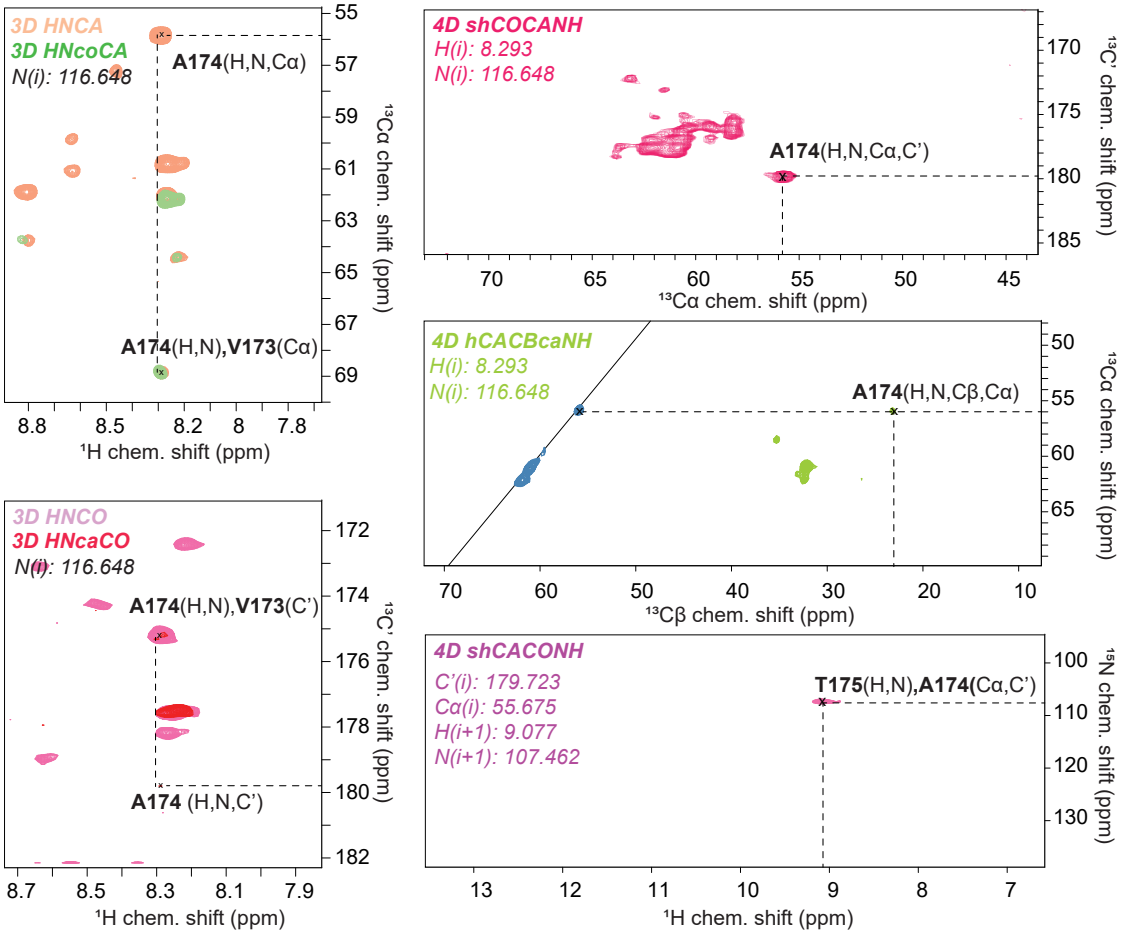


Figure 3.6: Hybrid assignment strategy applied to PfTrpB chemical shifts. Solution 3D HNCA/HNcoCA and 3D HNCO/HNcaCO spectra serve as starting points in the procedure. In the 3D HNCA spectrum, the H/N pair of Ala174 is correlated with its intraresidual $^{13}C_{\alpha}$ resonance. Same H/N coordinates in 3D HNCO/HNcaCO experiments give the ^{13}CO resonance of Ala174. Navigating to the determined H/N coordinates in 4D sehCOCANH experiment confirms $^{13}C_{\alpha}$ and ^{13}CO chemical shifts (and helps identify ^{13}CO chemical shifts in cases where peaks in solution HNcaCO experiment are absent), and in 4D hCACBcaNH experiment provides $^{13}C_{\beta}$ chemical shift. Navigating to the $^{13}C_{\alpha}$, ^{13}CO plane of 4D sehCACONH experiment one can see one (or more) peaks corresponding to Thr175. This peak will have identical $^{13}C_{\alpha}$ and ^{13}CO, but different ^{1}H and ^{15}N coordinates compared to Ala174. Navigating back to 3D HNCA/HNcoCA and 3D HNCO/HNcaCO experiments, backbone assignment continues until interresidual connectivity cannot be established unambiguously.

Assigned regions partially cover the COMM domain (Arg100–His190) and, more importantly, activation loop Val154-Thr160. In addition, catalytic residue Lys82 covalently bound to PLP is also within the assigned stretch of residues (Fig. 3.7). These assignments provide a foundation for residue-specific comparisons of activation-loop and active-site behavior between $PfTrpB^{wt}$ and $PfTrpB^{2B9}$, and for future ligand-dependent perturbation and dynamics measurements.

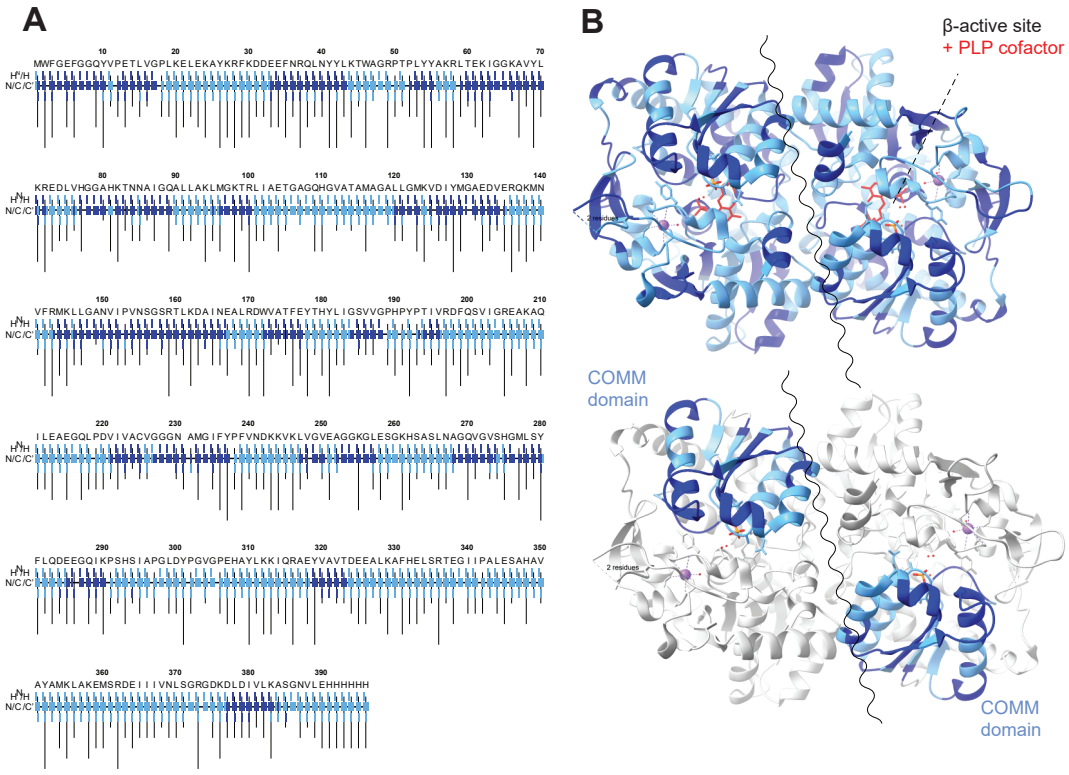


Figure 3.7: Partial assignments of PfTrpB represented in standard FLYA output format **(A)** and mapped onto the TrpB-TrpB dimer structure **(B)**. Dark blue rectangles in (A) stand for assignments with strong support. The row labeled $^{1}H_N/^{1}H_\alpha$ shows for each residue ^{1}H of the left and $^{1}H_\alpha$ on the right (here $^{1}H_\alpha$ chemical shifts are absent). The $^{15}N/^{13}C_\alpha/^{13}CO$ row shows for each residue the ^{15}N, $^{13}C_\alpha$, and ^{13}CO assignments from left to right. The row below the $^{15}N/^{13}C_\alpha/^{13}CO$ shows the side-chain assignments for residues. In (B), color scheme is the same as in (A): light blue residues comprise unassigned parts of the protein, dark-blue residues comprise assigned stretches. In (B) below only COMM domain is colored.

3.5.5 Data quality and consistency assessment framework

Assignment procedures require determining the most reliable experiments and understanding how they complement each other. The framework proposed here consists of (a) analysis of peak overlap, which quantifies the number of peaks visible within the chemical shift tolerance range assigned to each dimension; (b) analysis of assignment ambiguity, which calculates how many residues can be matched as $i-1 \rightarrow i$; and (c) evaluation of deviations between solution and MAS peak coordinates for the same residue, quantifying how chemical shifts of identical atoms from identical spin systems may differ between solution and MAS experiments. This pipeline establishes a hierarchy of experimental reliability and distinguishes coordinate differences that fall within experimental error from those representing genuine chemical shift perturbations or incorrect spin system assignments (Fig. 3.8).

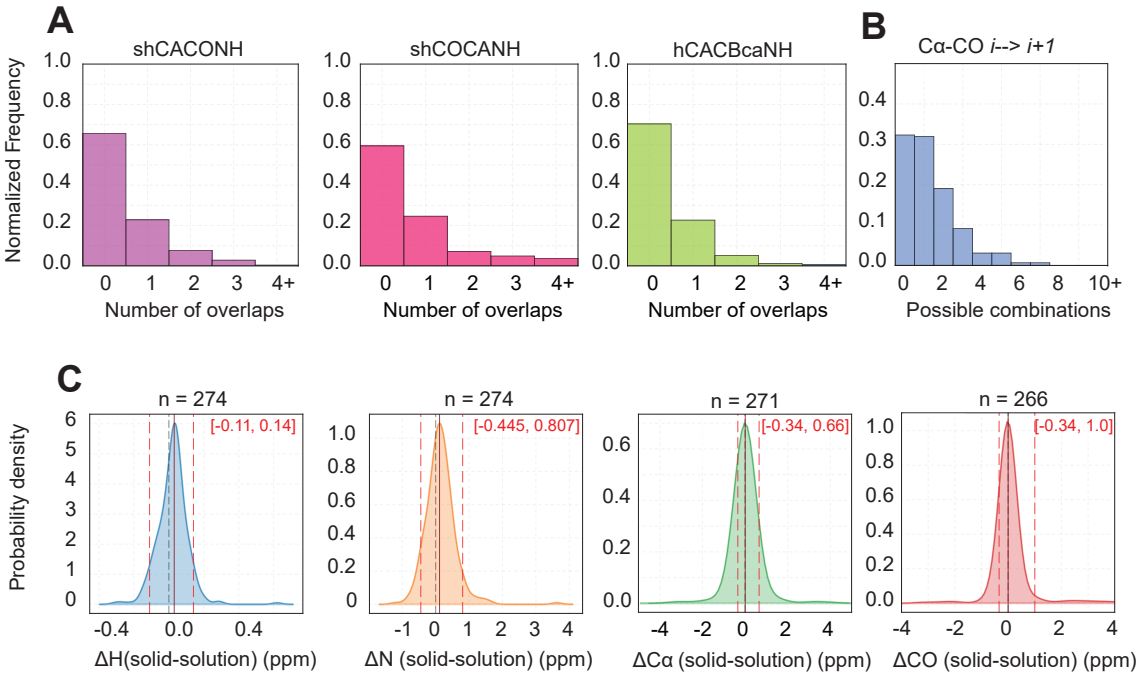


Figure 3.8: Statistical description of spectral quality and reliability. **(A)** Peak overlap analysis in 4D MAS NMR experiments. Overlap analysis of H/N correlated regions showing normalized frequency of spectral crowding. Peaks within tolerance ranges (0.075 ppm ^{1}H, 0.36 ppm ^{15}N) were grouped, and overlaps counted as additional peaks with distinct carbon coordinates. Numbers indicate overlaps per H/N region; 4+ represents four or more overlapping peaks. **(B)** Sequential assignment connectivity ambiguity analysis. Distribution of possible $i-1 \rightarrow i$ linkages for each spin system based on matching $^{13}C_{\alpha}$ and ^{13}CO chemical shifts between shCACONH and shCOCANH experiments. Spin systems were matched by H/N coordinates (tolerance: 0.075 ppm ^{1}H, 0.36 ppm ^{15}N), then potential sequential connections identified by $^{13}C_{\alpha}/^{13}$CO coordinate matching (tolerance: 0.29 ppm) with different H/N values. Values represent number of possible combinations per spin system. **(C)** Chemical shift deviations between solution- and MAS NMR spectra. These KDEs reflect only residues that could be matched between solution and MAS spectra; large CSP cases may be missing due to assignment/matching constraints. Solution NMR data were combined from HNCO, HNCA, HNcaCO, and HNcoCA experiments, MAS data used shCACONH and shCOCANH spectra. Red dashed lines indicate 90% confidence intervals, dark red solid lines show median values.

Since the combined dataset provides only $^{13}C_\alpha/^{13}CO$ connectivities, assignment success depends on spectral dispersion in these dimensions. A comprehensive list of chemical shifts of ^{1}H, ^{15}N, $^{13}C_\alpha$, ^{13}CO, and $^{13}C_\beta$ atoms for each peak was generated from a set of MAS NMR spectra shCOCANH, shCACONH, and hCACBcaNH. This list was then subjected to overlap analysis for H/N combination of nuclei using a custom Python script. The radius of overlap for ^{1}H and ^{15}N atoms was set according to previously established values: 0.075 ppm and 0.36 ppm for ^{1}H and ^{15}N, respectively [136]. Peak overlap analysis demonstrated favorable spectral dispersion across the three 4D experiments (Fig. 3.8, A). Most H/N correlated regions in each experiment contained single, non-overlapping peaks (shCACONH: ≈65%, shCOCANH: ≈55%, hCACBcaNH: ≈70%), indicating strong spectral resolution despite the inherent crowding challenges of large protein systems. Higher-order overlaps, defined as three or more peaks per region, were minimal in all experiments, facilitating reliable peak assignment and reducing ambiguity in sequential connectivity determination. These results indicate that the 4D experiments employed here provide sufficient resolution to ameliorate the overlap in three dimensions that is frequently observed in 3D hCONH and 3D hCANH experiments.

While peak overlap quantifies local spectral crowding, connectivity ambiguity quantifies how uniquely sequential links can be inferred even when peaks are resolved. The degree of ambiguity present in our MAS NMR dataset was assessed based on chemical shift lists assembled earlier for the overlap problem assessment (Fig. 3.8, B). Sequential assignment ambiguity was quantified by treating each spin system as a multidimensional vector defined by $^{1}H(i)$, $^{15}N(i)$, $^{13}C_\alpha(i)$, and $^{13}CO(i)$ chemical shift coordinates derived from complementary shCACONH and shCOCANH experiments. For each reference spin system i, potential sequential connectivities were identified by locating target spin systems whose $^{13}C_\alpha(i-1)$ and $^{13}CO(i-1)$ values matched the reference $^{13}C_\alpha(i)$ and $^{13}CO(i)$ coordinates within specified tolerance thresholds (0.29 ppm), while requiring divergent H/N chemical shifts to ensure distinct residue identity. The resulting ambiguity distribution, expressed as the normalized frequency of possible $i-1 \rightarrow i$ combinations per spin system, enabled a quantitative evaluation of backbone assignment confidence and the determinability of sequential connectivity within the experimental dataset. As a result, 35% of spin systems showed zero possible connections, consistent with (a) chemical exchange broadening, (b) immediate connection of the residue of interest to a proline residue, or (c) missing $^{13}C_\alpha/^{13}CO$ resonances due to tight tolerance thresholds. Approximately 30% of spin systems showed exactly one possible $(i-1) \rightarrow i$ connection, representing ideal unambiguous sequential assignments. Altogether, 0 and 1 possible combination groups account for more than 60% of peak connections, which is in line with previously published ambiguity analysis data for hCAII protein [136]. Both sensitivity-enhanced hCOCANH and hCACONH experiments give rise to only one or two possibilities for linking for about 50% of the resonances, and this value can likely be increased by adding 4D hCACBcaCONH or 4D HNcocaNH experiments to pair with hCACBcaNH and connect $^{13}C_\beta(i)$ and $^{13}C_\beta(i-1)$ residues in addition to already existing matching couples of $^{13}C_\alpha(i)$-$^{13}C_\alpha(i-1)$ and $^{13}CO(i)$-$^{13}CO(i-1)$.

Chemical shift perturbations between solution- and MAS conditions were assessed by constructing comprehensive spin system lists from complementary triple-resonance experiments, with solution-state data consolidated from HNCO, HNCA,

HNcaCO, and HNcoCA spectra and MAS parameters derived from shCACONH and shCOCANH datasets. Residue-specific chemical shift tensors were matched across experimental conditions based on assignment identity, and delta displacement scalar $\Delta = |\Delta\delta^{1}\mathrm{H}| + |\Delta\delta^{15}\mathrm{N}| + |\Delta\delta^{13}\mathrm{C}|$ were calculated for each nucleus, with statistical distributions visualized through 1D KDE plots to quantify the magnitude and directionality of chemical shift variations (Fig. 3.8, C). All nuclei display a normally distributed probability density with respect to their average chemical shift deviation between solution and solid-state spectra. Proton shifts ^{1}H showed minimal deviation of 0.01 ppm, nitrogen shifts ^{15}N displayed broader distributions with slight systematic differences. Carbon nuclei exhibited the largest dispersions, with $^{13}C_{\alpha}$ shifts spanning several ppm and ^{13}CO shifts showing both positive and negative deviations. These distributions define acceptable tolerances and identify outliers that need to be taken into consideration for the resonance assignment of PfTrpB.

3.6 Discussion

This study investigates and compares the structural and dynamic properties of protein samples prepared using various methodologies and experimental conditions. The objective was to determine how sample form (sedimented, crystalline, or solution), preparation methodology (batch crystallization and vapor diffusion crystallization), and the presence of certain stabilizers, such as glutaraldehyde at different concentrations influence spectral quality and stability. These factors ultimately determine the feasibility of samples for multidimensional 3D- and 4D NMR experiments in both solid-state and solution NMR settings.

Batch crystallization yielded significantly broader ^{1}H lines (median 189.5 Hz) than slowly grown vapor-diffusion microcrystals (≈82.6 Hz), indicating that microcrystalline slowly growing crystals provide the best trade-off of spectral resolution and SNR suitable for advanced NMR experiments. Additional stabilization of such crystals with various concentrations of glutaraldehyde results in significantly improved sample stability towards paramagnetic dopant Cu-EDTA in high concentrations, but induces severe line broadening in MAS spectra. Future work should explore milder and more selective cross-linking agents, and more tightly controlled reaction conditions to improve crystal stability without compromising spectral resolution.

The sedimented preparation displayed substantially larger time-dependent change in hydration than the microcrystalline sample, which may compromise long-term sample stability and contribute to the reduced sensitivity. At the same time, 1D ^{1}H spectra indicated 85% higher bulk signal intensity for the sediment relative to microcrystals which does not stem from different CP transfer efficiency or more favorable T_1 time for sediment. Instead, the intensity increase is most consistent with higher effective amount of protein due to higher rotor filling factor. Comparison of peak counts between three samples from hCONH and HNCO spectra yielded 286 peaks for the sedimented, 326 peaks for microcrystals, and 309 peaks for solution PfTrpB samples, although absolute counts depend on the manually chosen SNR threshold. Linewidth statistics further highlight nucleus-specific trade-offs: proton ^{1}H linewidths are comparable between the sediment and microcrystalline samples, whereas nitrogen ^{15}N and carbonyl ^{13}CO linewidths are narrower for the sediment.

Overall, sediments retain a more solution-like chemical environment but exhibit more pronounced time-dependent hydration drift and may suffer from lower sensitivity and sample instability during long experiments. Microcrystals are more stable in hydration and provide better sensitivity, but show larger perturbations relative to solution.

An optimized microcrystalline PfTrpB^{wt} sample enabled acquisition of the 4D MAS experiments shCACONH, shCOCANH, and hCACBcaNH. These spectra were used in combination with solution-derived 3D spectra for PfTrpB^{2B9} HNCO, HNCA, HNcoCA, HNcaCO, and HNCACB to obtain 39.8% of backbone resonance assignments, including parts of communication (COMM) domain, activation loop, and catalytic Lys82 residue. To evaluate the internal consistency and assignment utility of this combined dataset, we implemented a Python-based reliability framework comprising (a) analysis of $^{13}C_{\alpha}$-^{13}CO peak overlaps in H/N planes, (b) ambiguity of $(i-1) \rightarrow i$ connections, and (c) solution–MAS chemical-shift displacements for matched assignments. Overlap analysis showed that 65–70% of H/N regions contain a single peak and fewer than 20% contain two peaks, indicating that most local overlaps are limited and can often be resolved using solution-derived H/N anchors. Of all spin systems, 35% show zero possible $(i-1) \rightarrow i$ connections, whereas 30% yield a single unambiguous connection, representing the most assignment-favorable cases. Chemical shift dispersion was observed predominantly for $^{13}C_{\alpha}$ and ^{13}CO dimensions, consistent with increased sensitivity of carbon shifts to changes in sample state. These distributions provide an empirical basis for defining tolerances in FLYA and for flagging outliers that may reflect genuine perturbations or potential mis-assignments.

3.7 Limitations and future research

Several limitations should be noted. First, the current backbone resonance assignment coverage (39.8%) leaves substantial gaps that may include functionally important regions. Extending and validating the assignments will require additional complementary experiments that improve sequential connectivity and sidechain discrimination, for example a set of 4D HNcocaNH and 4D hCACBcacoNH experiments in solid-state and a 3D HNcoCACB experiment in solution NMR. Second, future comparisons between solution and MAS datasets should explicitly account for sample-dependent chemical shift displacements, particularly when using optimized slowly grown microcrystals that can deviate more strongly from solution reference conditions. Finally, while glutaraldehyde cross-linking increased crystal robustness against dopant exposure, it compromised spectral quality under the conditions tested. Exploring more mild and selective cross-linking agents may provide improved stabilization without introducing the line broadening observed here.

Chapter 4

cGAS N-terminal domain structure and dynamics

This chapter is based on work published in [137] and is reproduced here with modifications.

All MD simulations were done by Dr. Dmitrii Blokhin.

4.1 Introduction

To survive, organisms need the ability to recognize and fight invading pathogens. Two complementary types of immune response evolved in animals: innate immunity that relies on pattern recognition receptors, and adaptive immunity mediated by clonal expansion of antigen-specific lymphocytes. Different classes of pattern recognition receptors have evolved to recognize distinct pathogen-associated molecular patterns, one of which is cyclic GMP-AMP synthase (cGAS) [138]. cGAS is a Mg^{2+}- and Zn^{2+}-binding enzyme that belongs to the Mab21-like superfamily of nucleotidyltransferases (NTases) and is a major DNA sensor in humans and other animals. It is comprised of a hypervariable disordered N-terminal domain (NTD) and structured Mab21 domain, and recognizes both exogenous and endogenous double-stranded DNA (dsDNA), such as pathogenic DNA [139] and mitochondrial DNA [140] (Fig. 4.1).

cGAS exists as inactive monomers in the cytosol, which are activated by direct binding to dsDNA followed by dimerization and conformational rearrangements. Upon binding to dsDNA, cGAS undergoes conformational rearrangement and catalyzes the synthesis of the secondary messenger 2′,3′-cyclic guanosine monophosphate adenosine monophosphate (2′,3′-cGAMP) using ATP and GTP as substrates. This ligand 2′,3′-cGAMP then binds to the stimulator of interferon genes (STING) located on endoplasmic reticulum membrane to activate the expression of type I interferon (IFN-I) immune response [141].

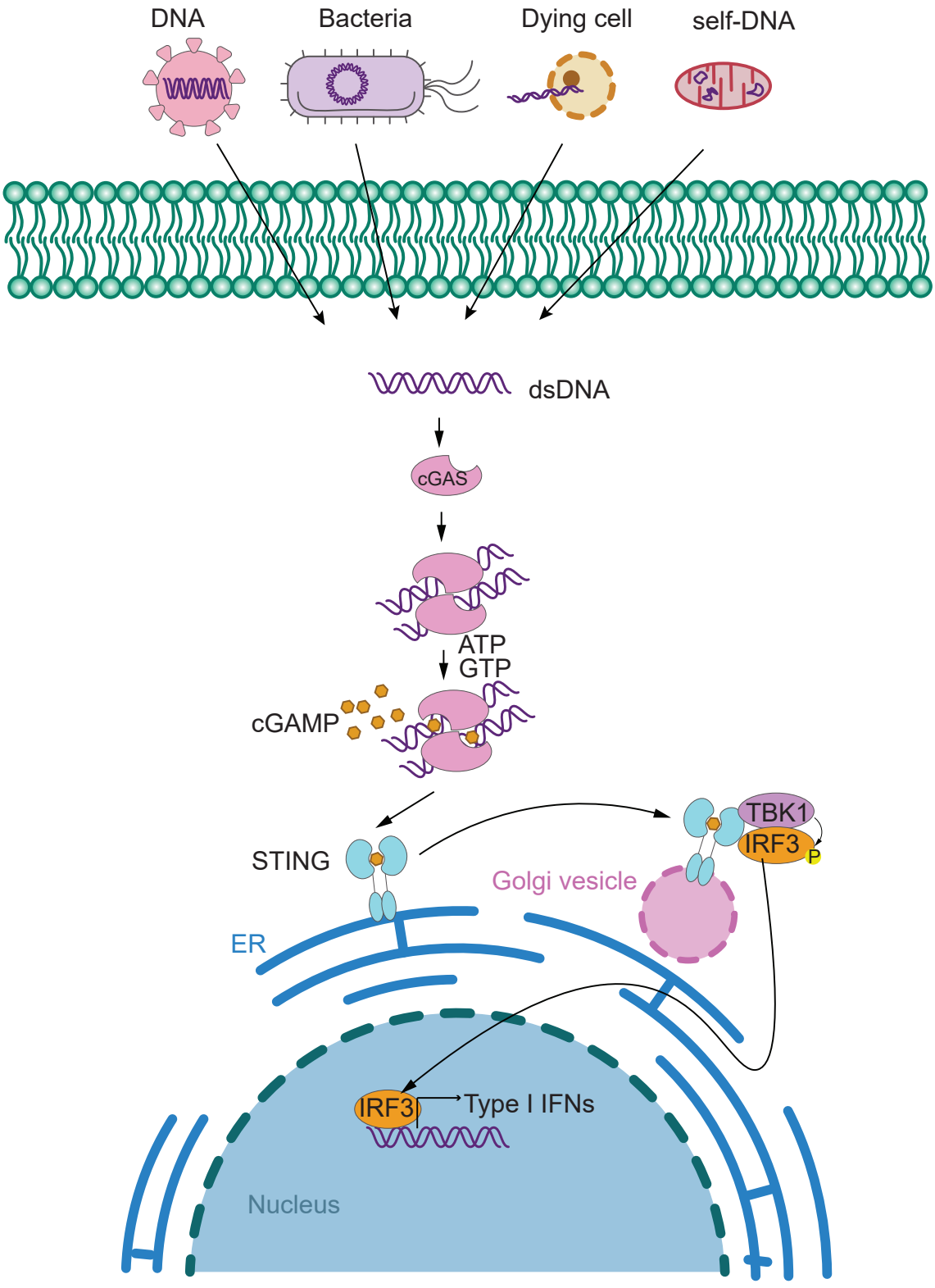


Figure 4.1: cGAS signaling cascade can be activated by dsDNA of various origin.

However, recent works challenge the generalized model of cGAS structure, function, and activity regulation. cGAS exists not only as a cytoplasmic but also as a nuclear protein [142][143] that is largely inactivated by interaction with nucleosomes [144][145] and chromatin DNA [146]. This prevents cGAS from constant up-regulation and limits autoimmunity. In the nucleus, cGAS has additional non-canonical functions: DNA damage repair [147][148], regulation of DNA replication [149], and control of chromosome fusion [150].

cGAS activation is predominantly sequence-independent, but cGAS is particularly enriched in satellite III repeats of centromeres and LINE1 elements on chromatin [151]. It requires DNA of at least 20 bp long for activation in cellular culture, yet in the presence of G-overhangs this length requirement can be lowered to 12 bp, and generally the presence of G-overhangs activates cGAS in a sequence-specific manner [152]. Human and murine variants of cGAS have different length specificity, where human cGAS requires DNA of at least 45 bp length for productive 2′,3′-cGAMP synthesis [153], while mouse cGAS is activated with 20 bp long DNA [152].

Despite being a dsDNA sensor, cGAS has been shown to interact with a small circular DNA cia-cGAS which prevents activation of IFN-I signaling in hematopoietic stem cells and thus protects them from IFN-I-induced exhaustion [154]. Typically, cGAS relies on DNA-mediated liquid-liquid phase separation (LLPS) to facilitate efficient cGAMP synthesis [155]. However, recent findings indicate that sequence-specific single-strand RNA (ssRNA) binding to cGAS serves as a driving force for condensate assembly, which subsequently enables the exchange of RNA for DNA to activate 2′,3′-cGAMP synthesis [156]. These findings demonstrate that cGAS activation is regulated by context-dependent biophysical factors, such as localization, nucleosome state, nucleic acid architecture, and condensate dynamics.

The main obstacle to obtaining a residue-resolved view of cGAS regulation is the presence of an intrinsically disordered NTD in cGAS. Usually, this domain is cleaved off for structural biology studies, and its functions and affinities were identified from biochemical assays [157][158] and mutation screenings [159]. Combined efforts of multiple groups resulted in a more comprehensive understanding of the role of NTD in cGAS function. Mechanically, NTD has been shown to play major role in LLPS and non-specific DNA-binding, helping cGAS to distinguish between self- and non-self DNA [159][160]. Tao et al. reported that nonspecific dsDNA binding by NTD induces conformational rearrangements, favoring the formation of transient α-helices. This refines the current model, which holds that the core alone, with its three DNA-binding sites (A, B, and C), is enough to trigger the STING/IFN-I pathway. Single-molecule imaging experiments revealed that NTD not only expands the range of motion of cGAS and helps the full-length protein to slide more readily along the λDNA, but also makes LLPS aggregates more fluid, allowing for substrate exchange [159].

Zn^{2+} ions have been shown to bind to human NTD in sites that overlap with DNA-binding regions and enhance the LLPS of cGAS NTD [161]. Because of the high net charge of NTD, it was suggested to interact with negatively charged phosphate backbone of DNA via a non-specific electrostatic interaction. This interaction can be further modulated by post-translational hyperphosphorylation/dephosphorylation cycles mediated by Aurora kinase B and cyclin-dependent kinase CDK1-cyclin B complex [145, 157, 138].

Kranzusch et al. proposed an additional role for NTD, observing that its presence significantly enhances the stability of cGAS proteins during purification. They linked this effect to stabilization or autoinhibition observed in other innate immune receptors [162][163]. Mechanisms of this particular stabilization effect are, however, unknown.

Despite numerous studies aimed at elucidating the role of NTD in cGAS catalysis, only two works have examined cGAS NTD at residue-specific level. Lee et al. explored how NTD enhances cGAS activity both *in vivo* and *in vitro* using analytical ultracentrifugation, circular dichroism (CD) spectroscopy, and NMR [160]. Jiang et al. investigated the role of NTD in LLPS of human cGAS employing various methods, including NMR [161]. In both studies authors showed that NTD displays secondary structure formation upon DNA binding, and identified the residues involved in this binding. Both studies provided backbone resonance assignments, which is a bottleneck for the majority of protein NMR studies, however, to our knowledge, these assignments were not deposited in the Biological Magnetic Resonance Data Bank (BMRB).

4.2 Aim of the project

This project investigates the structure and dynamics of the cGAS N-terminal domain using a combination of solution NMR and MD simulations to elucidate NTD's role in regulating cGAS activity. Initially, we sought to verify the assignments reported by Lee et al. [160] and subsequently use these assignments to characterize the structure and dynamics of cGAS NTD in both apo- and DNA-bound states. Based on existing literature, we hypothesized that cGAS exhibits a certain degree of sequence specificity, which we aimed to test using NMR titrations and experiments probing fast- and intermediate-timescale dynamics. These investigations provide insight into the fundamental mechanisms underlying cGAS activation and, more broadly, into how the internal dynamics of intrinsically disordered regions (IDRs) influences ligand binding.

4.3 Materials and Methods

4.3.1 Bioinformatics analysis of cGAS

DNA and protein sequences of cGAS orthologs were obtained from NCBI (https://www.ncbi.nlm.nih.gov/) and UniProt (https://www.uniprot.org/) databases using BLAST search (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [164][165]. Protein sequences of cGAS orthologs were identified by blastp search of the mouse cGAS full-length protein sequence against the Reference Proteins (refseq protein) database using the BLOSUM62 algorithm [166]. To retrieve more distant orthologs of cGAS from reptiles and amphibians, the chicken cGAS sequence was used. A list of 36 organisms was assembled to cover all clades of vertebrates with at least 3 species per family. Protein sequences were then uploaded to the Batch CD-Search engine (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) against the default CDD-62456 PSSMs database with an expect value cutoff of 0.0001 to obtain protein domain annotations and domain boundaries. Multiple sequence align-

ment (MSA) of NTD, Mab-21, and Mab-21 C domain sequences was prepared using the MUSCLE algorithm implemented in MEGA12 (https://www.megasoftware.net/) [167]. MSA of Mab-21 C domain sequences from all vertebrate species was further trimmed using Gblocks 0.91b (https://kbase.us/applist/apps/kb_gblocks/run_Gblocks/release) on relaxed settings to extract conserved sequence blocks [168]. The best-fit substitution model was selected using the Model Selection module in MEGA12 and used for further phylogenetic tree reconstruction. Phylogenetic trees of cGAS paralogs and orthologs spanning vertebrates, as well as an exemplary tree of cGAS paralogs in Mammalia, were reconstructed using Bayesian Evolutionary Analysis Sampling Trees (BEAST) (BEAST X) software (https://beast.community/) [169]. In each case, simulations were run under a strict clock model with a coalescent constant size tree prior in four parallel chains for 10,000,000 iterations, sampling every 5,000 iterations. A maximum clade credibility tree with nodes as median heights was generated using TreeAnnotator v10.5.0 with 100,000 burn-in states. To interpret inferred phylogenetic trees, the posterior probability calculated as the proportion of trees of a particular topology sampled during the MCMC run was used. Posterior probability values were considered to have strong support if within 0.95-1 interval, moderate-to-strong if within 0.9-0.95, and weak-to-moderate if within 0.7-0.9. The phylogenetic tree of selected species was obtained from NCBI Taxonomy Browser (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) [170] and optimized for better visualization in FigTree 1.4.4 (https://tree.bio.ed.ac.uk/software/figtree/). CD-Search-retrieved domain boundaries were mapped onto phylogenetic tree using a custom python script. Protein repeats were identified using RADAR tool (https://www.ebi.ac.uk/jdispatcher/pfa/radar) [171], and 3D structure of cGAS homologs was reconstructed using AlphaFold3 webserver (https://alphafoldserver.com/) [172]. Visualization of all 3D structures was performed in Chimera X19 (https://www.cgl.ucsf.edu/chimera/) [173].

4.3.2 Protein expression and purification

For cloning of NTD portion of cGAS we used a source plasmid vector pET28M-SUMO-mcGAS(full-length) kindly provided by Prof. Dr. Karl-Peter Hopfner laboratory in Genezentrum LMU Munich. Ligase-independent cloning, DNA sequencing, *E.coli* RIPL (DE3) pLysS transformation, test protein expression at 20°C, 25°C in autoinduction conditions, and 37°C followed by microcapillary SDS-PAGE to assess the protein amount were kindly performed by Tina Gazdag in the Protein Chemistry Facility of the Max Planck Institute of Molecular Physiology. The pOPIN-8xHis-eGFP plasmid was identified as the most effective vector for producing cGAS NTD (residues 1-146) from *M. musculus* (Supplementary Fig. B.4). The target protein was overexpressed in *E. coli* (DE3) RIPL cells as an eGFP fusion protein containing a PreScission (3C) protease cleavage site. For uniform ^{15}N and ^{13}C protein labeling, a single colony of transformed cells was inoculated into 1 L of standard M9 minimal medium supplemented with 100 μg/ml ampicillin, 1 g/L mathrm, 2 g/L ^{13}C-D-glucose, and 2 g/L ^{15}N/^{13}C-labeled ISOGRO® algal extract. Cells were cultivated at 37 °C and 110 rpm until an OD_{600} of 0.8 was reached. The culture was then chilled on ice for 20 min, and protein expression was induced by adding isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were

further incubated with shaking at 120 rpm, and 20 °C for another 20 h, followed by harvesting via centrifugation at 4 °C and 4,000 g for 20 min. Cell pellets were separated from media, and frozen at −80 °C until further use.

For purification, frozen cell pellets were thawed at room temperature and resuspended in **Q buffer** (See Supplementary Table A.4). After vortexing, cells were lysed using an Avestin Emulsiflex C3 French press. Lysates were aliquoted into 45 ml tubes and centrifuged at 100,000 g and 4 °C for 20 min, then incubated at 100 °C for 10 min to precipitate cellular proteases and centrifuged again as described above. The cleared lysate was applied onto a 5 ml HisTrap HP column (GE Healthcare) using an ÄKTA Pure chromatography system (GE Healthcare) in **Q buffer**. The eGFP-mNTD fusion protein was washed with 1 M NaCl and 1 M guanidine hydrochloride (Gu-HCl) to remove DNA, and eluted with a gradient of **U buffer** at 150 – 200 mM imidazole concentration. Elution fractions were then pooled together and concentrated to 1 - 2 ml using Amicon Ultra centrifugal filters with a molecular weight cut-off (MWCO) of 30 kDa (Sigma-Aldrich). The protein was buffer-exchanged into **3C buffer** using a PD10 desalting column (Cytiva). The protein was then mixed with 12 mg of 3C protease (prepared in-house) and cleavage was performed overnight at 4 °C on a rocking shaker. After cleavage, the protein mixture was applied onto a manually packed Ni-NTA gravity flow column (5 ml) equilibrated with **Q buffer**. A total volume of 70 ml of the flow-through fraction containing the protein was collected. The protein was then concentrated to a 2 ml volume and exchanged in **M buffer**. For NMR sample preparation, 270 μl of protein solution was mixed with 37 °C to a final concentration of 10% (v/v) and 0.1 M DSS to a final concentration of 1 – 3 mM. To analyze the quality of samples prepared, a standard SDS-PAGE with no modifications was used (see Chapter 2. Materials and Methods). Because the protein lacks aromatic amino acids, determining the exact concentration of the final sample is challenging. Other methods, such as the Bradford assay, also perform poorly with highly charged IDPs/IDRs [174]. We therefore reverse-calculated the concentration of the protein from the ratio of DSS and protein peaks in the 1D water-suppressed spectrum.

4.3.3 NMR data acquisition and processing

For the backbone NMR resonance assignment, the ^{1}H/^{15}N/^{13}C-labeled sample with the concentration of c ≈1 mM in 5 mm Shigemi NMR tube was used. ^{1}H/^{15}N-HSQC, ^{15}N/^{13}C-HSQC, 3D HNCO, 3D HNCA, 3D HNcoCA, 3D HNcaCO, 3D CBCANH, 4D HNcocaNH experiments were recorded. All NMR spectra were recorded at 25°C (unless specified otherwise) on a Bruker Avance NEO 800 MHz (^{1}H Larmor frequency) spectrometer equipped with a CP TCI proton-optimized triple resonance cryoprobe. The 4D HNcocaNH experiment [175] was recorded using non-uniform sampling (NUS) with 0.1% sampling density and reconstructed using the SSA algorithm [132]. Backbone resonance assignments of murine cGAS NTD have been deposited in BMRB (BMRB Entry 52649).

For the NMR relaxation measurements, fast timescale dynamics experiments (^{15}N hetNOE, longitudinal (T_1) and transverse (T_2) relaxation), as well as (^{15}N CPMG relaxation dispersion (RD) were recorded using standard pulse sequences [176]. ^{15}N CPMG-RD experiments on apo NTD and NTD bound to random42 DNA were carried out at 25°C and at ^{1}H frequency of 800 MHz, with the protein

concentration in apo NTD c = 1 mM, and in NTD + random42 DNA complex c = 66 μM. These experiments were measured in the same buffer as mentioned for the titration experiment. The constant time relaxation delay was set to T_{relax} = 0.066 s for apo NTD sample, and T_{relax} = 0.045 s for NTD + random42 DNA sample, and 14 CPMG frequencies were employed ranging from 100 to 2000 Hz.

For NMR titration experiments, nine DNA molecules were purchased from Eurofins Genomics, and added to the ^{1}H,^{15}N-labeled protein with c = 100 μM in a standard 5 mm NMR tube, together with 10 mM DSS (e.g., 10 μl of 0.1 M DSS stock solution) and 10% (v/v) D_2O. DNA sequences can be found in Supplementary Table A.9. CSP values were calculated for all residues by subtracting the chemical shifts of the proteins in the mixtures with DNA from the chemical shifts of the protein alone, using the following equation:

$$\text{CSP} = \sqrt{\frac{\left(\Delta\delta^{\,1\text{H}^\text{N}}\right)^2 + \left(0.1\,\Delta\delta^{\,15\text{N}}\right)^2}{2}}. \tag{4.1}$$

All molecular dynamics (MD) simulations were performed by Dr. Dmitrii Blokhin. Simulations were carried out in GROMACS [177] using the CHARMM36 force field for the protein [178] and the TIP3P water model [179]. The system was solvated in a cubic box (initial volume 10,648 nm^3; 407,532 water molecules) and neutralized by addition of 12 Cl^- ions. After steepest-descent energy minimization (maximum force threshold 1000 kJ mol^{-1} nm^{-1}) and short NVT equilibration (100 ps at 300 K using the Berendsen thermostat [180]), an additional water-vacuum slab setup was generated by removing half of the solvent to create a hydrophilic-hydrophobic interface, and simulated for 15 ns. The resulting structure was then used as the starting conformation for production simulations in a fully solvated periodic box, with equilibration under NPT conditions using the Parrinello-Rahman barostat [181], followed by a 1 ms production trajectory.

4.3.4 Data processing and plotting

NMR data were processed using TopSpin 4.5.0 (Bruker) (https://www.bruker.com/en/products-and-solutions/mr/nmr-software/topspin.html) and nmrPipe (NIST IBBR) (https://spin.niddk.nih.gov/bax/software/NMRPipe/NMRPipe.html) and analyzed using CCPNMR Analysis v. 3.2.4 (https://ccpn.ac.uk/)[110]. For processing and plotting, custom in-home Python scripts were used. Automated sequence assignment was performed in FLYA [134], and prediction of chemical shifts was performed in POTENCI [182]. Secondary structure propensity (SSP) was calculated using IDP-optimized algorithm CheSPI [183]. $R_{2,\text{eff}}$ and R_{ex}rates from CPMG experiments were obtained using customized Python script (courtesy of Dr. Dmitrii Blokhin) from peak intensities recorded at a series of CPMG refocusing frequencies (ν_{CPMG}). For each residue, intensity profiles $I(\nu_{\text{CPMG}})$ were converted to effective transverse relaxation rates $R_{2,\text{eff}}$ according to

$$R_2^{\text{eff}}(\nu) = -\frac{1}{T_{\text{relax}}} \ln\left[\frac{I(\nu)}{I_0}\right], \tag{4.2}$$

,where I_0 was defined as the maximum valid intensity across all measured ν_{CPMG} values for that residue. Dispersion curves $R_2^{\text{eff}}(\nu_{\text{CPMG}})$ were fitted (excluding the ν_{CPMG} = 0 point) either to a constant model (no exchange; $R_2^{\text{eff}} = R_2^0$) or to a

two-state fast-exchange Carver-Richards model (Supplementary Fig. B.5). DSS was used for direct ^{1}H chemical shift referencing, ^{13}C and ^{15}N chemical shifts were referenced indirectly.

4.4 Results

4.4.1 Bioinformatic and Evolutionary Analysis of cGAS NTD

Comparative phylogenetic analysis of Mab21-domain–containing proteins indicates that this family diversified early in animal evolution into several distinct clades corresponding to the Mab21-like family (Mab21L1, Mab21L2, Mab21L3, Mab21L4), Transmembrane protein 102 (TMEM102), MB21D2, Inositol 1,4,5-trisphosphate receptor-interacting protein family (IPTRIP, IPTRIPL1, IPTRIPL2), cGAS, and a group of transmembrane mitochondrial dynamics proteins (MID49 and MID51) (Fig. 4.2, A). Mab21-like proteins constitute a well-supported monophyletic group of Mab21L1, Mab21L2, Mab21L3, and Mab21L4 families (Fig. 4.2, B) which share a common evolutionary origin and, to some extent, developmental functions: Mab21L1 and Mab21L2 regulate embryonic eye development [184, 185, 186, 187], Mab21L3 establishes the dorsoventral axis in the developing epidermis [188][189], and Mab21L4 mediates transforming growth factor-β (TGF-β) signaling in differentiated epidermal keratinocytes [190]. However, the activating ligands of Mab21-like proteins and their potential secondary messengers remain largely uncharacterized. MB21D2 and TMEM102 proteins form a distinct clade, yet their known functions differ significantly: MB21D2 modulates cGAS-STING signaling in myeloid cells [191], whereas TMEM102 regulates mitochondrial outer membrane permeabilization [192][193][194]. TMEM102 is not the only mitochondria-associated member of Mab21 superfamily: an unusual example of functional divergence is provided by MID49 and MID51. Both proteins are located in the outer mitochondrial membrane and form a complex with Dynamin-related protein 1 (Drp1) to promote mitochondrial fission, a process integral to the normal life cycle of mitochondria. [195][196]. Notably, cGAS has been shown to recapitulate certain aspects of MID proteins' functionality, e.g. localize in mitochondrial outer membrane of cancerous cells due to the presence of two cryptic mitochondria localization signals, and interact with Drp1 to protect cells from ferroptosis and mitochondrial ROS accumulation [197]. The observed functional convergence of MID proteins and cGAS, mediated by NTase domain, suggests a potential mitochondrial function of other Mab21-domain containing proteins.

The NTD of various cGAS orthologs shares an evolutionary origin with a number of herpesviral proteins, primarily tegument proteins (Fig. 4.2, C). In mammals, NTD is homologous to DNA polymerase subunits gamma and tau family, a fragment of UL36 large tegument protein, VP22 tegument protein, ICP4 transcriptional regulator protein, Epstein-Barr early antigen 3B (EBNA-3B), g150 protein from Epstein-Barr virus (BDLF3), FhaB hemagglutinin protein from toxin-antitoxin system of *Bortedella pertussis*. A more detailed look at reptilian, amphibian, and piscine cGAS NTD annotations reveals additional, less common fusions with low confidence, such as Merozoite erythrocyte binding protein 2 from *Plasmodium* (MAEBL), 60S ribosomal L19-like protein, transcription termination factor Rho, and pneumococcal surface protein PspC (Supplementary Table A.8). As an outgroup for all vertebrates,

a cGAS-like protein from *Branchiostoma floridae* was used, which possesses a provisional, low-probability adenylate kinase domain at its N-terminus. In many species, no identifiable NTD can be detected; however, this region consistently demonstrates a high predicted disorder propensity. Notably, vertebrate species with highly active innate immune systems, such as the naked mole rat (*Heterocephalus glaber*), several reptiles (*Alligator mississippiensis, Crocodylus porosus, Pantherophis guttatus, Crotalus tigris*), and amphibians (*Phyllobates terribilis, Xenopus laevis, Xenopus tropicalis, Ambystoma mexicanum*), lack any annotated NTD in their cGAS homologs.

These patterns suggest that (a) NTDs may have independently originated multiple times from an annotated domain that subsequently lost its characteristic sequence motifs through extensive diversification, and (b) N-terminal viral-like insertions are not essential for efficient innate immune activation and may instead represent an adaptation associated with immune modulation.

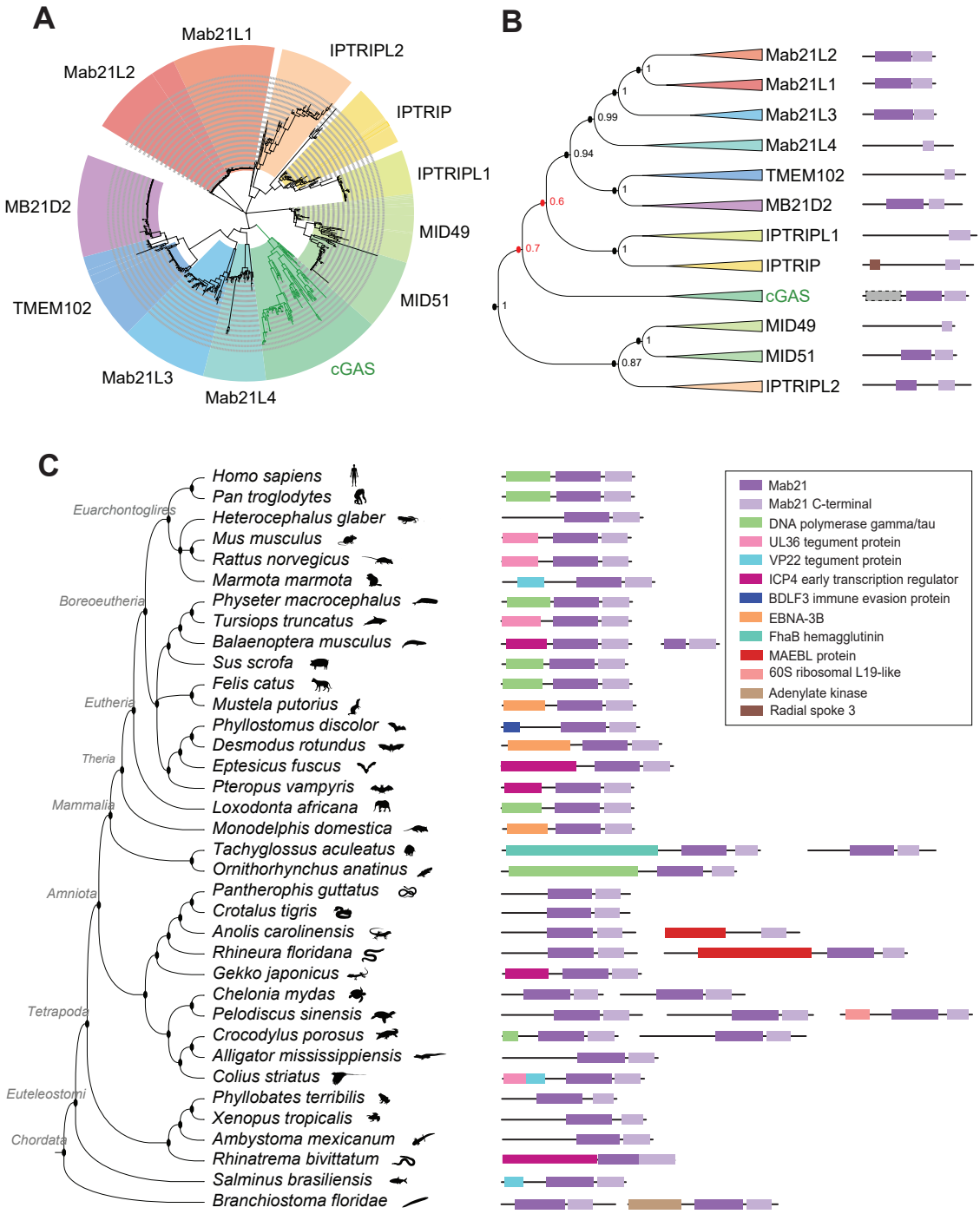


Figure 4.2: Evolutionary history of Mab21-domain containing protein family. **(A)**. Large unrooted tree of all Mab21 proteins. Branch lengths correspond to the number of amino acid substitutions. **(B)** Reconstructed evolutionary relations between various paralogous clades reconstructed from proteins from the following primate species: *Homo sapiens*, *Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla gorilla*, *Pongo abelii*, *Macaca mulatta*. Colors of individual families are identical to colors assigned to families in (A). Domain organization of an exemplary member of each protein family is mapped onto the domain evolution tree. Human orthologs were used as exemplary members. Mab21-domain is divided into N-terminal portion (dark purple) and C-terminal portion (light purple) according to CD-Search annotation. **(C)** Tree of species of selected vertebrates with domain boundaries and annotations retrieved using CD-search.

Analysis of the genetic diversity of NTDs from various animal species revealed multiple instances of tandem duplications averaging 13 to 20 amino acid residues in length. In monotremes, the NTD adopts a β-helical structure characteristic of FhaB β-prism domain, and shares a proportion of identical protein motifs. Each motif consists of a repeat unit and exhibits high mutability, which complicates the identification of identical consensus motifs between cGAS orthologs in platypus and echidna. Additional segmental duplications were identified in (a) a legless lizard *Rhineura*, where repeats also fold into a β-helix but display a distinct primary structure compared to monotremes, and (b) a legless amphibian *Rhinatrema*, where repeats instead form disordered loops. Furthermore, the NTD of cGAS of *Rhinatrema* folds into a helical domain of unknown annotation and function (Fig. 4.3). These findings indicate that tandem repeat expansion of cGAS NTD represents a characteristic feature of cGAS evolution across vertebrate species.

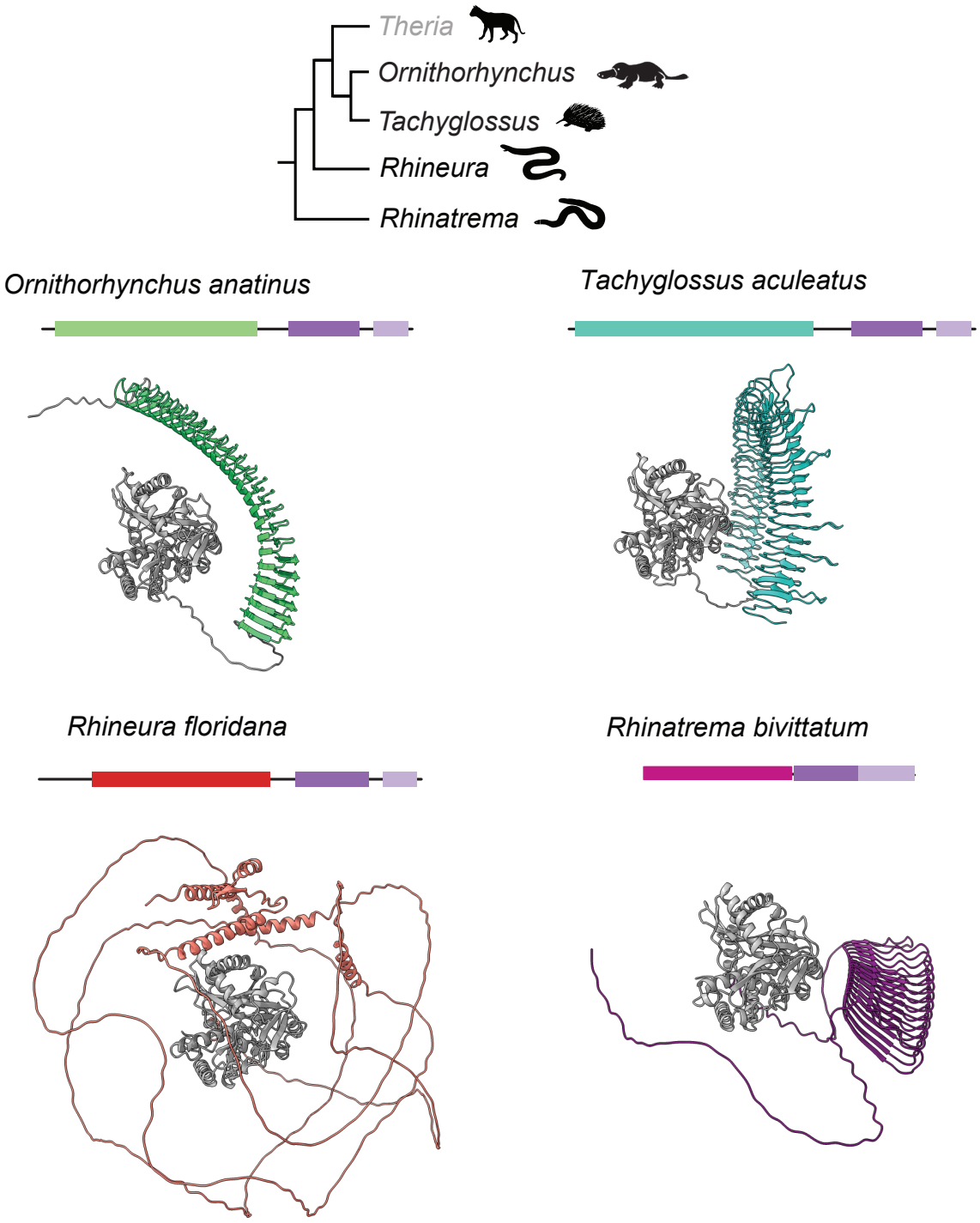


Figure 4.3: Examples of segmental duplication expansion events in NTDs of cGAS of platypus (*Ornithorhynchus anatinus*), echidna (*Tachyglossus aculeatus*), Florida worm lizard (*Rhineura floridiana*), and two-lined caecilian (*Rhinatrema bivittatum*) identified using RADAR and PRIGSA and visualized using AlphaFold3 and Chimera X19.

Although NTDs from different organisms are sometimes annotated as belonging to various domain families, a BLAST-derived set of NTD sequences from *Chiroptera*, *Rodentia*, *Primates*, *Carnivora*, *Lagomorpha*, *Artiodactyla*, and *Perissodactyla* orders share several conserved or semi-conserved residues, enabling MSA and quantitative assessment of trends in NTD evolution. Comparative analysis of cGAS sequences from rodents and primates reveals that the most prevalent conserved motifs are positively charged R/K-rich regions and hydrophobic P/A/L-rich repeats (Supplementary Fig. B.3). These features suggest that a combination of electrostatic attraction and hydrophobic forces drives the multivalent protein-protein and protein-ligand interactions. However, further experimental validation is required to confirm this hypothesis.

4.4.2 Protein production

For experimental studies of the cGAS NTD, the mouse ortholog was selected over the human variant due to its previously reported higher stability and activity [198]. The vector pOPINE-6xHis-eGFP-3C-mNTD demonstrated the highest protein yield ($\succ$250 ng/μl) in test expression (see Supplementary Fig. B.4) and was therefore employed for large-scale protein production. Codon-optimized cell lines that compensate for rare proline codons, such as *E.coli* RIPL (DE3) pLysS, are essential for efficient cGAS protein expression because of the high proline content in NTD (23 Pro/146 total residues). During purification, all fractions from the cell pellet to the eGFP-mNTD fusion sample exhibited an intense green color attributable to eGFP tag. The presence of green coloration in the cell debris pellet after the initial sonication and ultracentrifugation suggests that a portion of the protein is insoluble or membrane-associated, thereby reducing the effective yield in the final preparation. Partial recovery of the protein from the pellet can be achieved using CHAPS- or Triton-X100-containing lysis buffers; however, this approach did not result in a significant increase in the final NTD sample amount.

Following separation of the supernatant containing eGFP-mNTD from insoluble cell debris, the supernatant is incubated at 100°C for 5-10 minutes to denature most cellular proteases and protein contaminants; this method is commonly employed in the purification of thermophilic proteins [130]. A subsequent ultracentrifugation step yields a white pellet of denatured proteins, and a green supernatant containing eGFP-mNTD fusion. eGFP is not adversely affected by high temperatures, as it is an engineered protein with high temperature stability up to 75°C for extended periods and can tolerate even higher temperatures for shorter durations while maintaining its native fold [199]. Since NTD is a disordered protein, it is similarly unaffected by elevated temperatures and does not require refolding after heat treatment.

For Immobilized Metal Affinity Chromatography (IMAC), either Ni-NTA or cobalt-containing TALON resins can be used with a varying degree of efficiency: Ni-NTA resin typically yields higher quantities of protein, whereas TALON resin results in greater protein purity. In both protocols, inclusion of a reducing agent (β-mercaptoethanol or TCEP) in essential. Without a reducing agent, the protein irreversibly adheres to the resin beads and cannot be eluted even with high concentrations of imidazole or low pH buffers. The origin of the disulfide bonds responsible for this strong binding to the resin, whether from the eGFP or NTD regions of the fusion protein, remains unclear.

The presence of even little amounts (20-30 mM) of imidazole in binding or washing buffers consistently causes protein smearing and elution from resin. Incorporating high salt (300-500 mM NaCl) in the buffer is more effective for minimizing non-specific binding.

An additional critical step in NTD purification occurs during protein immobilization on the resin, where bound plasmid DNA is removed from the sample using 1 M NaCl and 1 M Gu-HCl. Eluting the eGFP-mNTD fusion directly from resin without this step can complicate protein concentration measurements, as the 260 nm absorbance from residual DNA can obscure the 280 nm protein signal. Furthermore, DNA that remains associated with the protein after proteolysis can obstruct the Amicon® membrane during subsequent protein concentration steps.

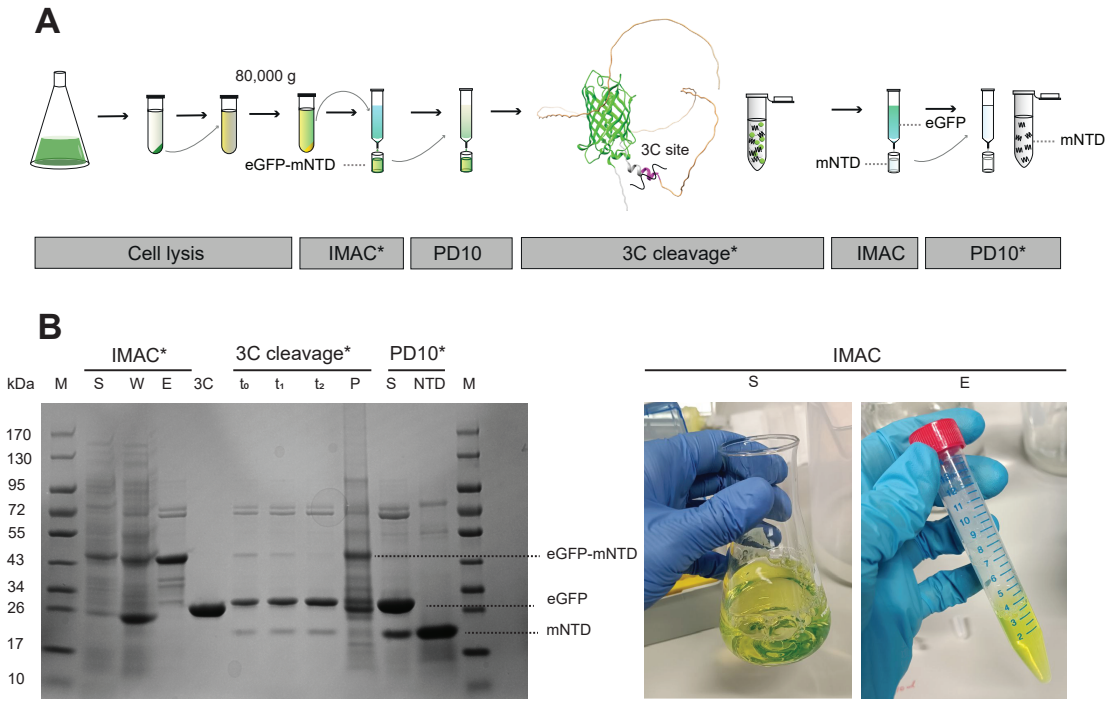


Figure 4.4: Production of mouse cGAS NTD. **(A)** Schematic of NTD purification from 1 L of $^{1}H/^{15}N/^{13}C$-labeled M9 media supplemented with 2 g/L culture $^{1}H/^{15}N/^{13}C$-labeled ISOGRO® growth media supplement. Details of expression and purification procedures are described in 4.3 Materials and Methods. High electrostatic attraction between eGFP and mNTD, as well as between mNTD and bacterial plasmid DNA required that the IMAC column with bound 6xHis-eGFP-mNTD is washed with 1 M NaCl and 1 M Gu-HCl prior to elution with imidazole-containing buffer. Resultant protein sample concentration is usually within 1.5-1.9 mM/ L of expression range. **(B)** SDS-PAGE gel to assess the procedure and purity of a final sample. M – Marker, S – Supernatant, W – Wash, E – combined Elution fraction, t0/t1/t2 samples correspond to the procedure of 3C cleavage of 6xHis-eGFP from mNTD using 3C Protease, P – Pellet (insoluble fraction after overnight proteolytic cleavage), S – Supernatant fraction after overnight proteolytic cleavage, NTD – final NTD sample concentrated in NMR buffer.

Following Ni-NTA purification, the protein sample should be desalted using either a PD10 column or dialysis. The PD10 desalting column is preferred for transferring the protein into a 3C protease buffer due to its robustness and efficiency (Fig. 4.4, A). Protein concentration at the eGFP-mNTD fusion stage can be performed using Amicon® devices with a molecular weight cut-off (MWCO) of 30K or 50K. For eGFP tag cleavage, in-house-produced 6xHis-HRV3C protease (3C protease) at a concentration of 1–2 mg/ml was used at a 1:1 molar ratio. This protease specifically recognizes the short peptide L-E-V-L-F-N-G-P at 4°C and cleaves between the N and G amino acid residues. After overnight cleavage, the protein sample typically separates into two fractions. The insoluble fraction consists of uncleaved eGFP-mNTD fusion, denatured 3C protease, and cleaved eGFP tag. The soluble fraction contains the majority of mNTD, soluble 3C protease, and trace impurities (see Fig. 4.4, B). To remove 3C protease, reverse Ni-NTA chromatography is performed so that the soluble eGFP tag, uncleaved eGFP-mNTD fusion, and 3C protease remain bound to the resin, while mNTD is collected in the flow-through fraction. To minimize the risk of mNTD co-eluting with the eGFP tag due to electrostatic interactions, it is recommended to include 1 M guanidine hydrochloride (Gu-HCl) in the buffer used for reverse Ni-NTA. This denaturant is subsequently removed during PD10 desalting or during protein concentration if the sample is diluted two or three times with the final NMR buffer. For final mNTD sample concentration, the MWCO of Amicon® devices should not exceed 10 kDa. In cases where eGFP contaminates the mNTD sample, a two-step concentration protocol can be employed. The protein is first concentrated with a 50 kDa MWCO device to separate uncleaved eGFP-mNTD fusion and eGFP dimers, followed by concentration of the flow-through fraction with a 10 kDa MWCO device. However, this two-stage separation is highly inefficient and results in a substantial portion of mNTD remaining in the concentrated heterogenous mixture of proteins during the first concentration step, likely due to electrostatic attraction between the highly positively charged mNTD (pI = 11) and the negatively charged eGFP (pI = 5.6).

Assessing the concentration of the final mNTD sample is challenging because the protein lacks aromatic amino acids and does not have absorption at 280 nm. Alternative protein quantification methods, such as the Bradford assay, also yield unreliable results, as this assay is unsuitable for precise concentration measurements in IDPs or IDRs due to their unique and variable amino acid compositions [174]. Hence, the mNTD concentration is routinely determined using a 1D ^{1}H-NMR experiment with water suppression, which enables calculation of the concentration of any ^{1}H-containing compound in a sample, provided a reference chemical such as DSS is present at a known concentration.

4.4.3 Backbone assignment and structural propensity

The mNTD ^{1}H/^{15}N-HSQC spectrum displays limited peak dispersion, with ^{1}H chemical shifts clustered between 7.7 and 8.7 ppm, which is indicative of a predominantly disordered structure (Fig. 4.5 A, B). Initial ^{1}H-detected 3D experiments suffered from signal overlap in the carbon dimension, making unambiguous assignments challenging. We therefore recorded a 4D HNcocaNH experiment [175] which benefits from the long T_2 time of IDPs and allows to obtain unambiguous sequential assignments in a reasonably short time. Automated sequence assignment using

FLYA achieved 77% assignments of non-proline residues, which was subsequently improved to 92% through manual assignment in conjunction with chemical shift prediction from POTENCI. Backbone amide resonances were assigned for all residues except Met1, Glu2, Asp3, Gln111, His127, Arg128, and Arg138. The protein primary sequence contains two repeats with the amino acid motif RGARS at positions 39–43 and 129–133. These repeats produce overlapping signals for Ala41 and Ala131 as well as for Arg42 and Arg132, which complicates unambiguous assignment. The number of assignments for individual types of nuclei is as follows: ^{1}H: 113, ^{15}N: 113, $^{13}C_{\alpha}$: 131, $^{13}C_{\beta}$: 120, and ^{13}CO: 134. The protein sequence comprises 146 amino acids, including 10 Gly residues (lacking $^{13}C_{\beta}$) and 23 Pro residues (lacking ^{1}H and yielding no ^{15}N assignments in triple-resonance experiments). As a result, 92% of backbone ^{15}N and ^{1}H resonances were assigned and. Excluding proline residues, 89% of $^{13}C_{\alpha}$, 96% of $^{13}C_{\beta}$, and 91% of ^{13}CO resonances were assigned.

$^{13}C_{\alpha}$ and $^{13}C_{\beta}$ NMR chemical shifts were used to assess deviations from random-coil values, facilitating the identification of potential α-helices, β-sheets, and disordered loops. CheSPI analysis indicated similar structural propensities, suggesting that the protein primarily adopts a random-coil conformation with short, transient α-helices near residues 115 and 120 (Fig. 4.5, C).

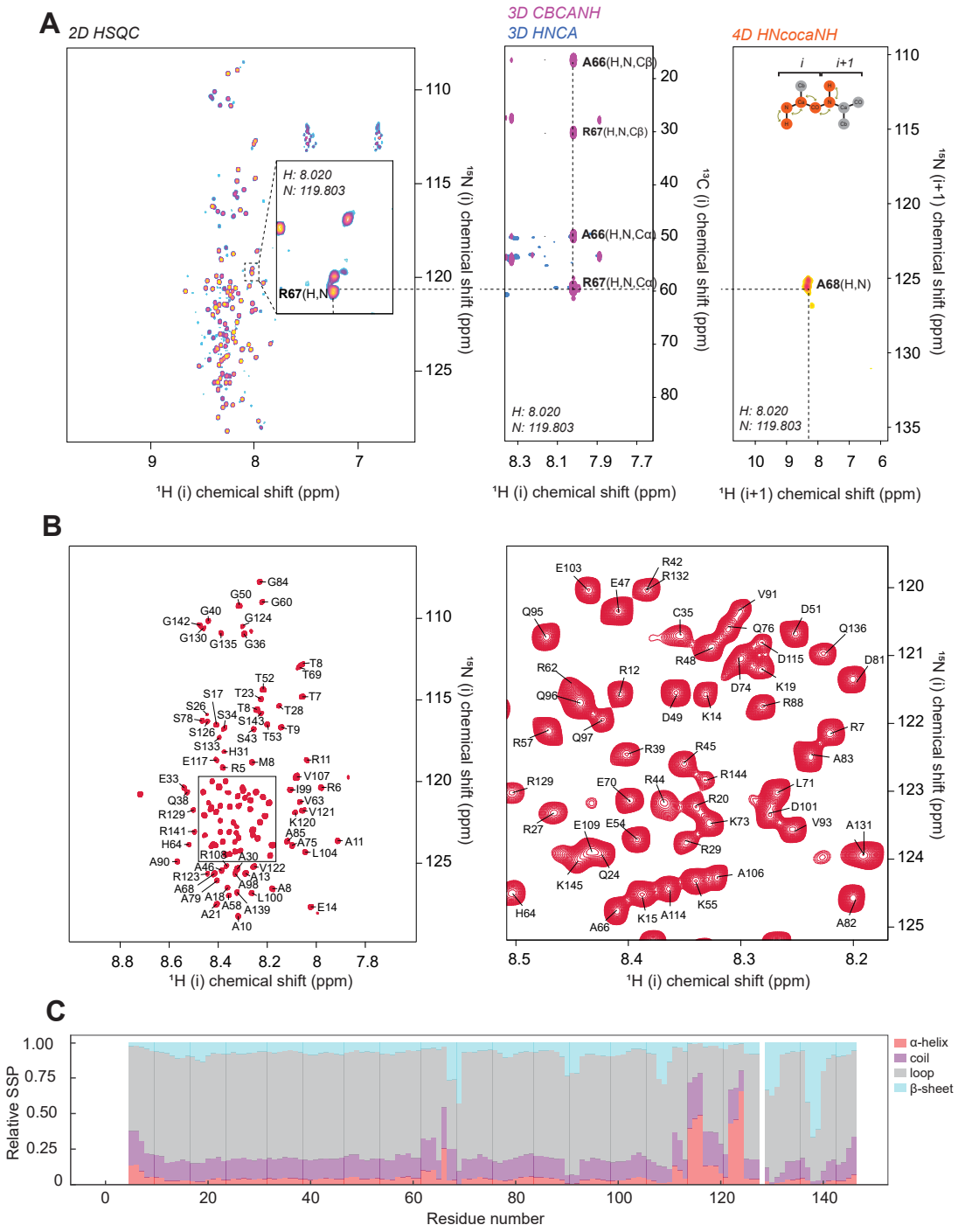


Figure 4.5: Backbone resonance assignment of a mNTD. **(A)** Assignment procedure using the 2D HSQC, 3D HNCO, 3D HNcaCO, 3D HNCA, 3D HNcoCA, 3D CBCANH, and 4D HNcocaNH experiments. The process starts with choosing a well-separated peak in the 2D HSQC followed by navigating to the 3D HNCA/CBCANH window to identify the amino acid types of residues i and $(i+1)$ based on $^{13}C_\beta$ and $^{13}C_\alpha$ chemical shifts. Navigating from the 2D HSQC to the 4D HNcocaNH window will usually result in seeing exactly one peak corresponding to amino acid (i+1). The connectivity can then be confirmed by comparison of $^{13}C_\alpha$, $^{13}C_\beta$, and ^{13}CO chemical shifts between complementary 3D experiments. **(B)** Assigned 2D ^{1}H,^{15}N-HSQC of mouse cGAS NTD, recorded at 800 MHz ^{1}H Larmor frequency at 25 °C. **(C)** Secondary structural propensity (SSP) of mouse cGAS NTD.

Proline cis/trans isomerization is a significant contributor to the conformational heterogeneity of IDPs. Determining the cis/trans proline ratio is essential, as distinct conformational states may underlie different biological functions. ^{1}H-detected NMR experiments do not detect proline residues, resulting in gaps in sequential backbone resonance assignments and requiring additional ^{15}N- or ^{13}C-detected experiments. To address the challenge of proline assignment, a dedicated ^{13}C-detected experiment (^{1}H/^{13}C-HSQC) was recorded. In combination with a 2D ^{15}N-^{13}CO projection of a 3D HNcaCO experiment, this approach enabled the assignment of ^{15}N chemical shifts of 13 proline residues (Fig. 4.6, A, B).

Analysis of $^{13}C_{\beta}(i-1)$ chemical shifts from 3D CBCANH experiment and ^{13}CO($i-1$) chemical shifts from the 3D HNCO experiment indicated that no additional peaks suggestive of cis-proline conformations were present. Based on the obtained S/N ratio, major secondary cis-proline populations can therefore be excluded. The assigned proline shifts all exhibit a $^{13}C_{\beta}$ chemical shift of approximately 32 ppm (Fig. 4.6, C) which is characteristic of trans-proline and is consistent with chemical shifts predicted by ncIDP [200]. However, minor cis-proline populations remain possible.

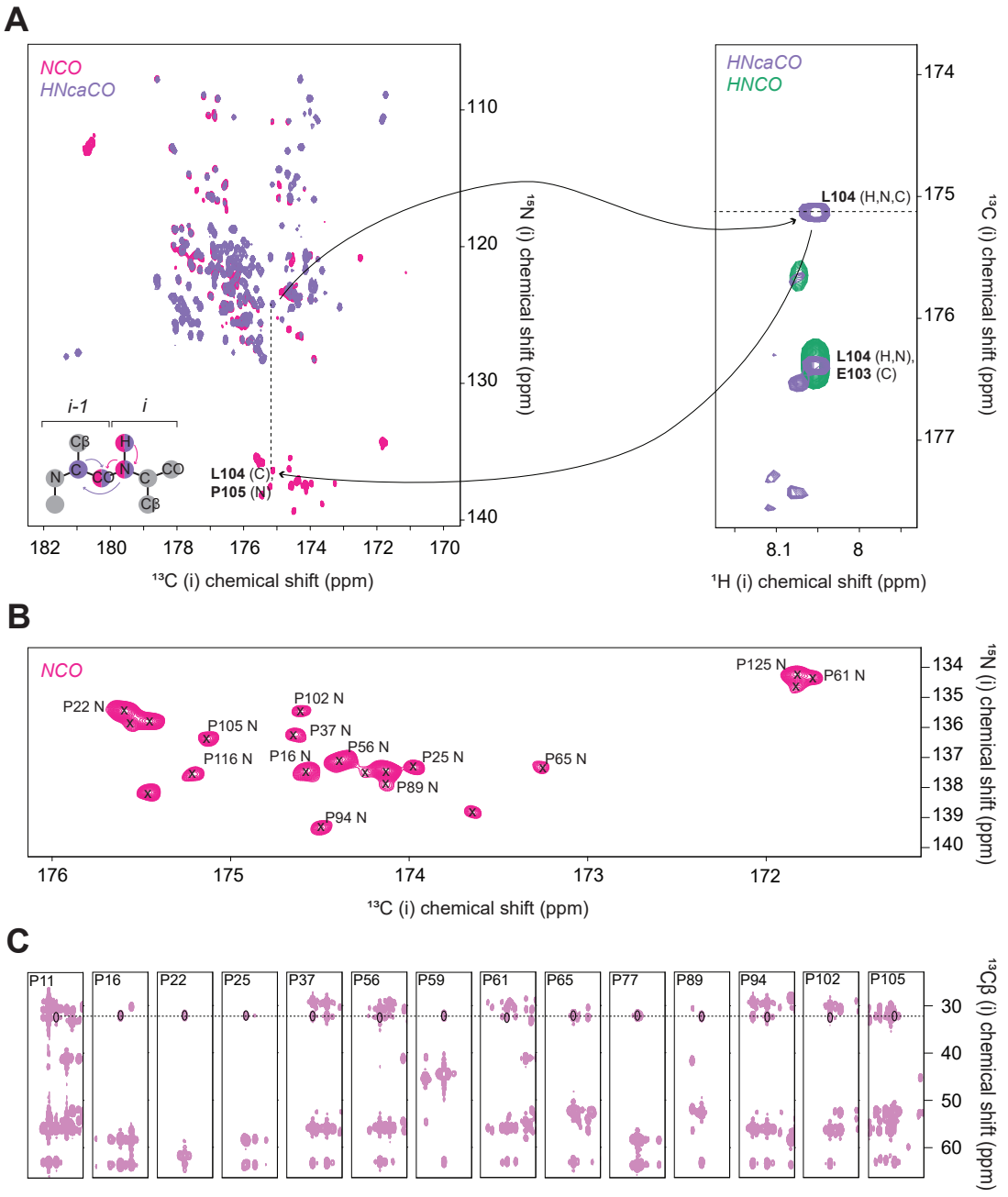


Figure 4.6: Proline content, partial backbone resonance assignment, and $^{13}C_\beta$ chemical shifts of selected proline residues of murine cGAS NTD. **(A)** $^{1}H/^{13}C$-HSQC spectrum (pink) and ^{15}N-^{13}CO projection of 3D HNcaCO experiment (purple) used for ^{15}N resonance assignment of proline residues. **(B)** Partial ^{15}N resonance assignment of 13 proline residues out of 21 peaks present and 23 peaks expected. **(C)** ^{1}H-$^{13}C_\beta$ slices at different ^{15}N coordinates obtained from a 3D CBCANH experiment to visualize $^{13}C_\beta(i-1)$ resonances grouping around 32 ppm corresponding to trans-proline conformation.

4.4.4 Fast timescale dynamics of apo NTD

The fast timescale dynamics of the apo NTD was characterized via ^{15}N R_1 and R_2 relaxation rates measured at 15°C and 25°C (Fig. 4.7, B). The relaxation profiles display features typical of IDPs, such as elevated R_1, moderate and heterogenous R_2 values, and residue-specific variability throughout the sequence. Analysis of individual relaxation rates, together with the pseudo-correlation time (pseudo-$\tau_c = R_2/R_1$), reveals a clear modular organization of NTD, with three regions displaying distinct relaxation signatures: an N-terminal segment (residues 1–45), an intermediate region (46–99), and a C-terminal segment (100–146) (Fig. 4.7, C). The N-terminal segment exhibits consistently elevated R_2 values and correspondingly higher pseudo-τ_c values, indicating locally slower motions and increased rigidity relative to the rest of the chain. These patterns suggest the presence of transiently ordered elements or restricted conformational backbone sampling in this region. In contrast, the intermediate region displays uniformly reduced R_2 and pseudo-τ_c values, consistent with enhanced ps–ns mobility characteristic of a highly flexible linker-like region. The C-terminal segment, again, shows higher R_2 and pseudo-τ_c values, comparable to those of the N-terminus, indicating that both termini sample more restricted dynamical states than the central portion of the protein. Collectively, the combined R_1/R_2/pseudo-τ_c profiles delineate three dynamically distinct subdomains within the apo NTD, each characterized by a unique balance of local flexibility and rigidity.

Mapping the pseudo-τ_c values onto the most populated conformational cluster obtained from the 1-μs MD simulation (Fig. 4.7, D) provides additional structural context for the modular organization of the NTD. Residues exhibiting elevated pseudo-τ_c values, indicative of reduced local mobility, are primarily clustered in the central region of the folded globule, particularly within the short helix spanning residues 30–35 and in a small rigid patch near residues 99–101. Additional isolated rigid sites are observed at the periphery (e.g., Gly60, His64, Ala90, Gly136), suggesting that local ordering is not uniformly distributed but is instead confined to discrete structural elements. The RMSF values from the same MD cluster (Fig. 4.7, E) support these conclusions: the central helix demonstrates significantly reduced fluctuations, while the C-terminal tail (residues 120–125 and 142–146) is identified as the most flexible region of the protein, showing the highest atomic mobility within the dominant MD conformational cluster (42.1% of sampled frames).

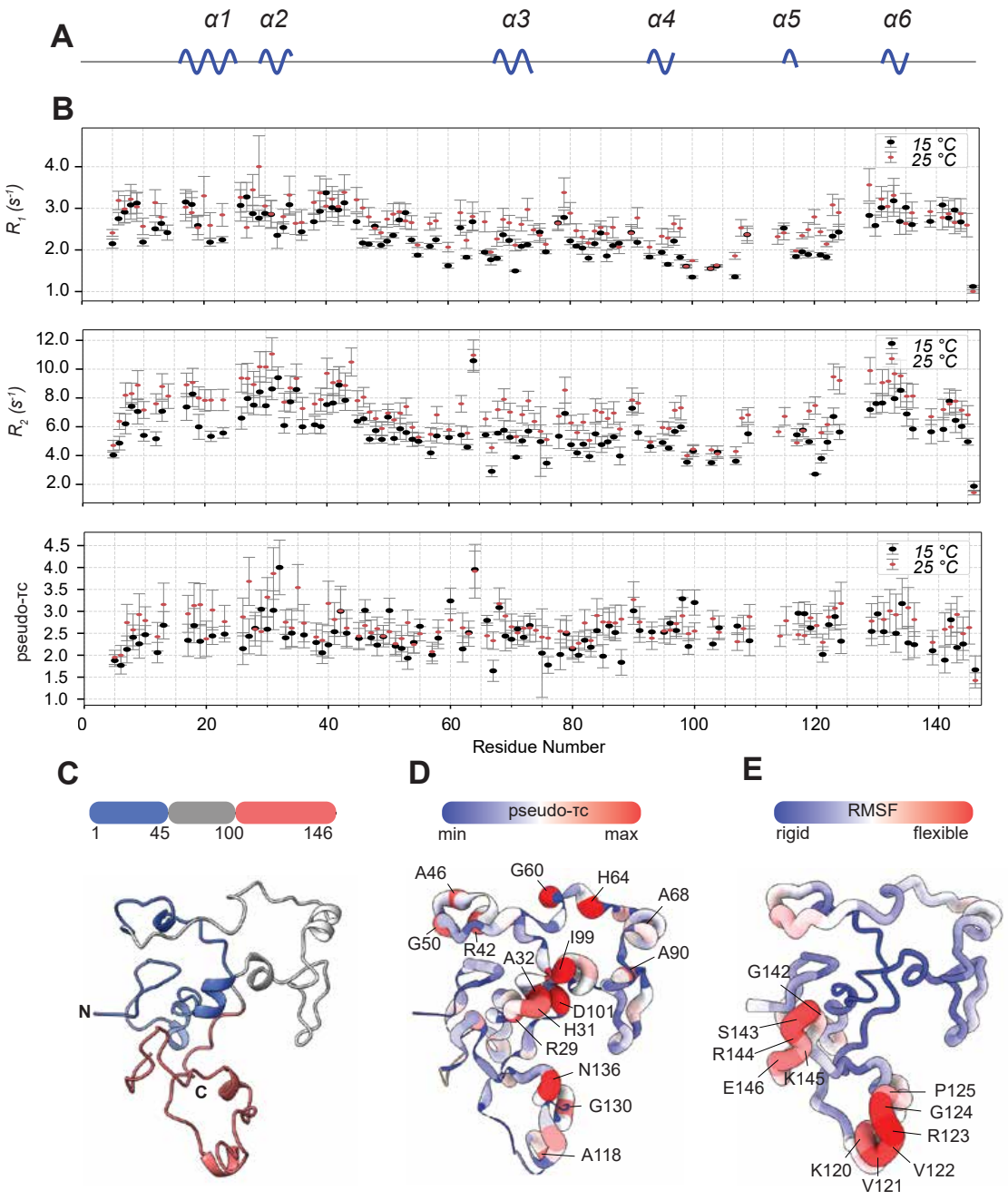


Figure 4.7: Fast timescale dynamics of the apo NTD. **(A)** Helical propensity of apo NTD derived from the most populated conformational cluster (42.1% of frames) from a 1-μs MD simulation. **(B)** Residue-specific R_1 and R_2 relaxation rates of apo NTD measured at 15°C and 25°C, and the corresponding pseudo-τ_c calculated as R_2/R_1 ratio that serves as a measure of a local rigidity. **(C)** Subdomain organization of NTD residues 1–45, 46–99, and 100–146) inferred from R_1, R_2, and pseudo-τ_c profiles mapped onto the MD-derived structure. **(D)** pseudo-τ_c values calculated for the 25°C dataset mapped onto the MD-derived structure. Higher pseudo-τ_c values indicate locally restricted backbone motion and frequently coincide with transient secondary-structure elements. **(E)** Root-mean-square fluctuation (RMSF) values of the same MD cluster, reporting on positional flexibility. Red regions are highly flexible; blue regions indicate restricted motion. Note that in D, red and blue refer to less and more motion, respectively, whereas in D, more motion is red, and less motion is blue.

This alignment between NMR-derived rigidity and MD-derived structural propensities (a) supports the division of the NTD into three dynamically distinct subdomains, and (b) demonstrates that the NTD samples a conformational landscape characterized by relatively rigid N- and C-terminal regions, interspersed with patches of temperature-sensitive residues, and connected by a flexible intermediate segment.

4.4.5 DNA-binding properties

NTD was shown to have an important role in the DNA-binding capacity of the full-length cGAS [160], but the region of interaction was not previously identified. To assess NTD interaction with dsDNA, and determine whether the NTD interaction pattern depends on DNA length, we compared binding to DNA duplexes ranging from 15 to 57 bp. The presence of equimolar amounts of dsDNA induced notable CSPs in the $^{1}H/^{15}N$-HSQC spectra of NTD (Fig. 4.8, A, B), with no sequence-dependent differences detectable under these conditions. The largest perturbations were observed for residues RRRT (residues 5-8) in the N-terminal segment and RKV (residues 119-121) plus Gly135 in the C-terminal segment. These residues are spatially proximal as follows from MD simulation, and serve as DNA interaction hotspots for all tested DNA molecules (Fig. 4.8, C, E). Together, the positively charged regions RRRT (residues 5-8) and RKV (residues 119-121) comprise the charged core of the DNA-binding motif. This motif bears certain resemblance to AT-hook motifs that are widely distributed in chromatin remodeling proteins and bind to the minor groove of AT-rich DNA, thus facilitating DNA bending and binding of other chromatin modulators [201]. Whether the NTD also induces DNA bending requires further corroboration.

The stacked I/I_0 profile for NTD–DNA interactions serves as a binding reactivity map across the sequence. All tested DNA produce reproducible intensity loss in the N-terminal (1–45) and C-terminal (100–146) regions, indicating DNA-induced line broadening and confirming these segments as the primary interaction regions (Fig. 4.8, D, F). In contrast, residues 66–99 show minimal changes in intensity, demonstrating that the intermediate region remains dynamically decoupled and does not participate in binding. This architecture (two binding-competent terminal modules separated by a flexible linker) is characteristic of IDRs with short linear motifs, enabling independent motion of terminal segments and supporting multiple binding modes [202][203].

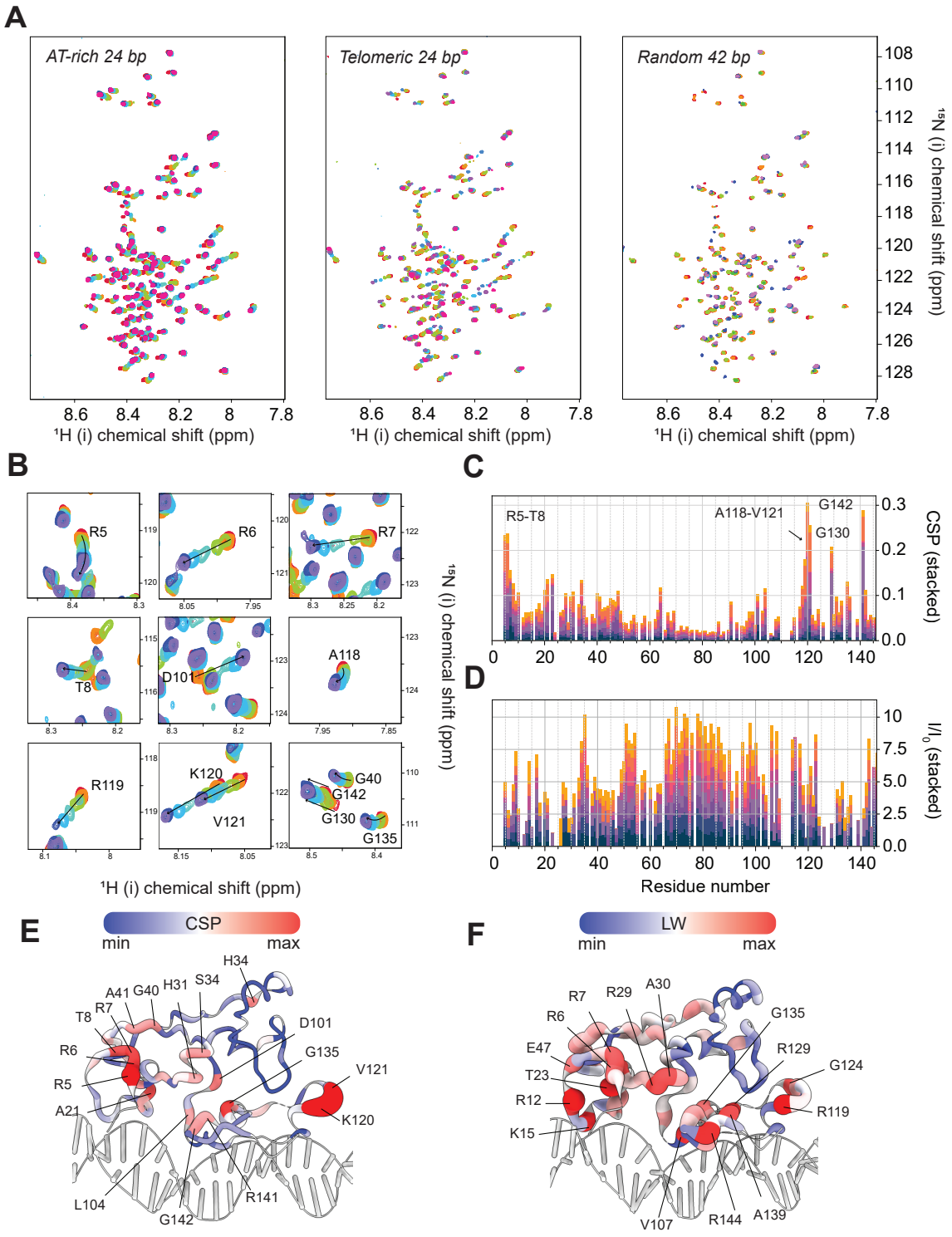


Figure 4.8: Multivalent interactions of cGAS NTD with various DNA molecules. **(A)** Overlay of ^{1}H/^{15}N-HSQC spectra of NTD during the process of titration with AT-rich 24 bp DNA, telomeric 24 bp DNA, and random 42 bp DNA. **(B)** Residues of NTD that experienced notable CSPs upon titration with ISM DNA. **(C)** Cumulative CSPs of NTD upon titration with seven different DNAs. **(D)** Cumulative peak intensity ratios from titration spectra of NTD alone (I_0) and after adding various DNA to the 1:1 molar ratio (I). **(E)** Cumulative CSPs mapped onto the structure of the most populated cluster from 1 μs MD simulation of NTD bound to HSV DNA (57 bp). **(F)** Cumulative I/I_0 values mapped onto the structure of the most populated cluster.

In the past, Tao et al. utilized CD spectroscopy to show that NTD displays signs of transient secondary structures formation upon DNA binding [159]. We therefore aimed to test this by making use of high sensitivity of chemical shifts to secondary structures. As a test DNA we chose herpesvirus-specific DNA (HSV DNA, 57 bp) and recorded 3D HNCO, 3D HNCA, and 3D CBCANH experiments to obtain ^{1}H, ^{15}N, ^{13}CO, $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ chemical shifts. Following SSP analysis using CheSPI showed that NTD-DNA complex has high probability of $\alpha 3$ helix (64-73 aa) formation, while $\alpha 5$-helix (118-122 aa) likely disappears (Fig. 4.9, E, F). Interestingly, $\alpha 3$-helix is located in intermediate region of NTD that does not interact with DNA according to CSPs and I/I_0 intensity analyses. This helix therefore may be fortified to serve different purpose.

Our MD simulations of NTD, NTD bound to DNA, and full-length cGAS revealed that NTD in a full-length protein is tethered to the cGAS NTase core via a stable α-helix, which we further identified as $\alpha 3$-helix (Fig. 4.9, A-C). The role of this helix therefore seems to be rather complex: on one hand, it does not interact directly with DNA, yet binding to DNA induces its formation; on another, it serves as a tethering anchor in full-length cGAS that realigns all three regions of NTD (and, therefore, alters the mutual arrangement of DNA-binding regions within NTD). To clarify the role of the $\alpha 3$-helix helix, we devised two alternative protocols: (a) titration of NTD with cGAS core followed by assessment of CSPs, SSP, as well as fast timescale dynamics, and (b) purification of a full-length cGAS followed by resonance assignment tranfer between NTD-only and NTD within cGAS proteins. Unfortunately, initial attempts to titrate NTD with core resulted in cGAS core precipitation due to buffer incompatibility, and purification of a full-length cGAS did not yield quantities enough for recording backbone resonance assignment experiments. Further work in optimization of buffer composition for cGAS NTD and core, as well as expression and purification protocols for a full-length cGAS would be needed.

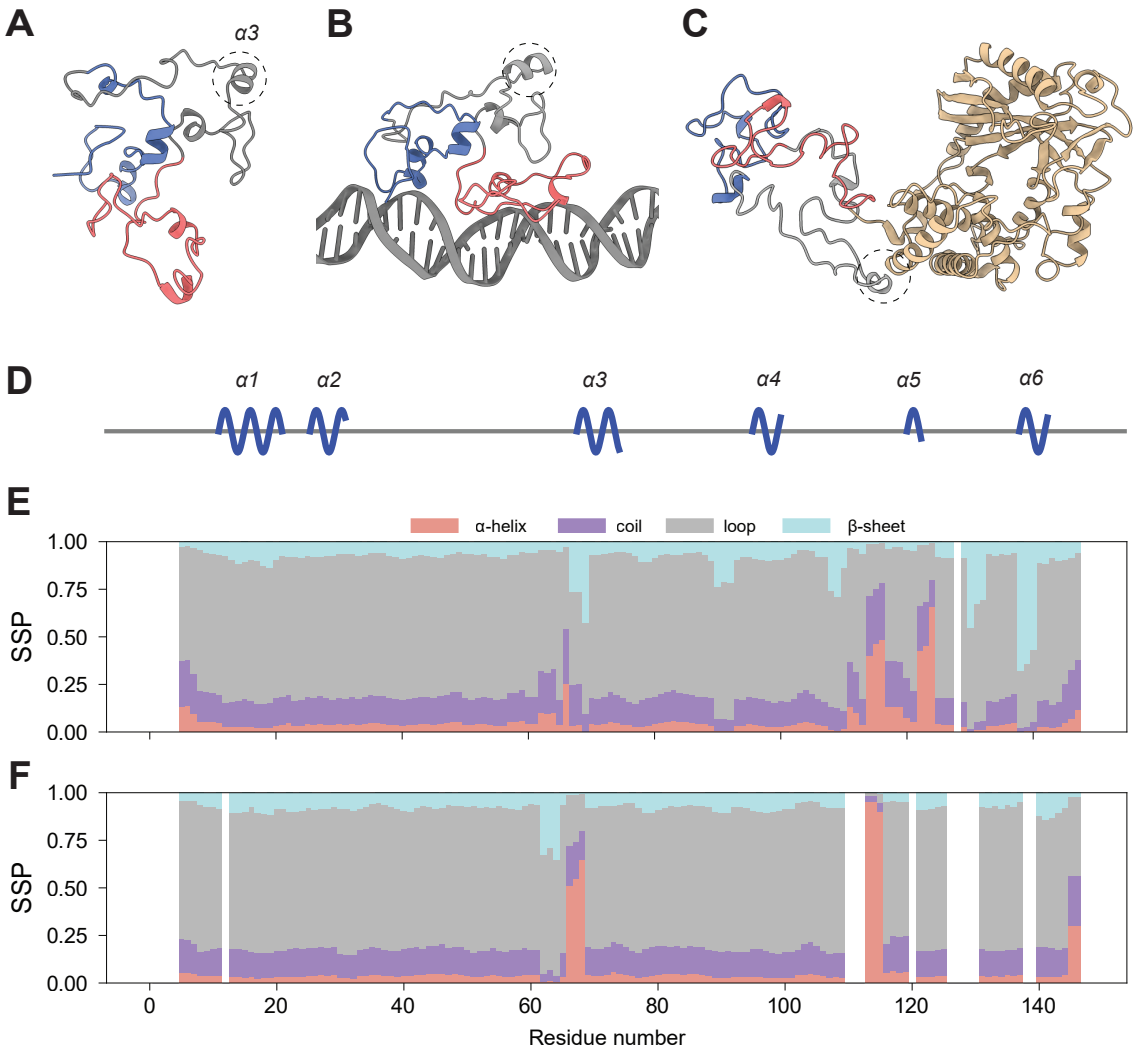


Figure 4.9: Hypothetical tethering of NTD to cGAS core via $\alpha 3$ helix (residues 64-73) detected via MD simulation. **(A)** Average structure of the most populated cluster (%) of apo NTD conformations with structural subdomains mapped. **(B)** Average structure of the most populated cluster (%) of NTD bound to HSV DNA with structural domains mapped. **(C)** Average structure of the most populated cluster (%) of full length cGAS with structural domains of NTD mapped. **(D)** Secondary structure elements in NTD bound to HSV DNA detected with MD simulation. **(E)** Secondary structure probability (SSP) in apo NTD calculated from NMR chemical shifts using CheSPI. **(F)** SSP in NTD bound to 1x HSV DNA calculated from NMR chemical shifts.

Comparison of HSQC titrations with dsDNA of similar length (20–24 bp) revealed two regimes: in some cases, resonances exhibited primarily chemical shift perturbations with minimal line broadening, while in others most peaks broadened beyond detection (Fig. 4.10, A). This observation suggests intermediate exchange dynamics (i.e., most probably association/dissociation exchange) driven by a DNA-dependent parameter other than length, in particular GC content. Using the number of broadened peaks as a primary measure of the amount of intermediate exchange, an approximately linear dependence on GC content was observed with moderate statistical support: DNAs with high GC content (e.g., rigid DNA) show significant line broadening compared to flexible DNAs with low GC content (Fig. 4.10, B). Because all duplexes were matched in length (20–24 bp), the observed differences in line broadening cannot be attributed to the duplex length (as can be seen from Fig. 4.10, C). This dependence indicates that DNA composition modulates the kinetics of the interaction, with higher GC DNA driving slower exchange (e.g., slower on and/or off rates) and more extensive broadening.

To quantify intermediate-exchange contributions more rigorously, ^{15}N CPMG relaxation-dispersion experiments were conducted on apo NTD and NTD bound to random42 DNA which represents a DNA that induced the most line broadening in entire set (42 bp, 72%GC) (Fig. 4.10, D). Reliable spectra and fits were obtained only for apo NTD, whereas CPMG data for DNA complex displayed inadequate fitting (Supplementary Fig. B.5) is likely due to (a) oversaturation of protein with DNA (2:1 ratio of protein to DNA), and (b) low protein concentration in protein-DNA complex. Future CPMG measurements will focus on DNA of similar length, and employ lower DNA concentration and higher protein concentration to extract R_{ex} values and enable comparison to apo NTD.

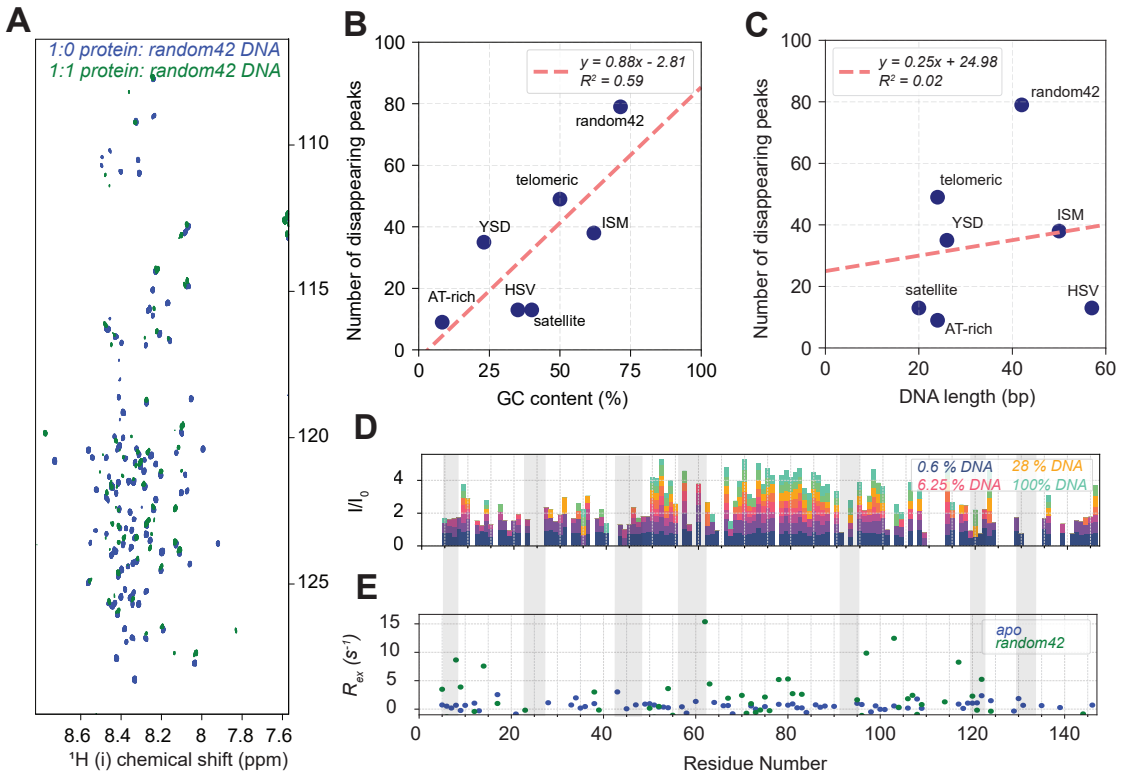


Figure 4.10: Difference in dynamic response of NTD to DNA of similar length but various GC content. **(A)** Comparison of response of NTD to random42 DNA (42 bp, 71.43% GC). **(B)** Linear dependence of number of peaks 'disappearing' upon DNA binding from the GC-content of added DNA. **(C)** Linear dependence of number of peaks 'disappearing' upon DNA binding from the DNA length. **(D)** Intensity profile (I/I_0) of NTD obtained from the NMR signal intensities in the $^{1}H/^{15}N$ HSQC spectrum of the free state of NTD and NTD bound to random42 DNA with different molar ratios of DNA: 20%, 45 %, and 100 %. **(E)** ^{15}N conformational exchange contributions, R_{ex}, extracted from CPMG relaxation dispersion data acquired at 800 MHz and 25°C on NTD. $c(NTD)_{apo} = 1$ mM, $c(NTD)_{NTD\text{-}DNA} = 66\ \mu M$.

4.5 Discussion

This study addressed an important element in understanding cGAS regulation: its intrinsically disordered hypervariable N-terminal domain (NTD). For decades, NTD has been excluded from structural analyses, resulting in limited residue-level characterization despite its established roles in DNA binding and liquid–liquid phase separation. Here, a combination of solution NMR, molecular dynamics (MD) simulations, and evolutionary reconstructions was employed to analyze structure, dynamics, and DNA-binding properties of murine cGAS NTD as a first step towards comprehensive characterization of the role of NTD in cGAS.

Bayesian phylogenetic reconstruction of an evolutionary history of cGAS NTD unexpectedly revealed that NTD likely shares evolutionary origin with tegument and regulatory proteins from *Herpesviridae*, a large and evolutionary successful family of double-stranded DNA viruses that have been coevolving with vertebrates for up to 400 million years [204]. Among these homologous viral proteins were: fragments of UL36 large tegument protein, VP22 tegument protein, ICP4 transcriptional regulator, Epstein-Barr early antigen 3B (EBNA-3B), and Epstein-Barr BDLF3 protein.

Possible viral origin of NTD may put cGAS under the partial physiological control of herpesviruses, which could have a profound effect on cGAS functionality, ranging from activation to attenuation of its antiviral properties. This hypothesis is indirectly supported by multiple reports on cGAS inhibition upon interaction with herpesviral proteins, such as direct inhibition of cGAS and cGAS-driven LLPS by VP22 tegument protein [205][206], and deamidation and following deactivation of cGAS by UL37 in UL36/UL37 complex [207].

The HSQC dispersion and relaxation behavior of the NTD exhibits the characteristic hallmarks of an IDP, with narrow ^{1}H chemical shift dispersion consistent with an overall random-coil ensemble interrupted by short transient helices. Upon combining a 4D HNcocaNH experiment with standard backbone resonance assignment experiments, we obtained a near-complete backbone coverage (92% of amide ^{1}H/^{15}N). Addressing proline cis-trans isomerization, we assigned 13 of 23 expected proline peaks, and observed no evidence for cis-proline conformers presence consistent with predominantly trans-proline chemical shifts predicted for IDPs, although minor cis-populations cannot be excluded.

The relaxation data discovered the modular organization of NTD, where the backbone can be divided into three regions with distinct relaxation signatures: N-terminal region, intermediate region, and C-terminal region. Titration of NTD with double-stranded DNA induced notable chemical shift perturbations in N- and C-termini, in particular in positively charged segments RRRT (5-8 aa) and RKV (119-121 aa). At the same time, a significant amount of line-broadening in these subdomains was also observed as the DNA concentration increased. We concluded that N- and C-termini of the NTD comprise the DNA-binding surface and are separated by a dynamically decoupled intermediate region that serves as a flexible linker. MD simulations further confirmed the folding of N- and C-termini into a DNA-binding surface and also revealed the formation of a stable α3-helix (64-73 aa) in the intermediate linker. This helix formed upon the DNA titration of NTD in NMR data, and is also involved in the tethering of the NTD to the C-terminus of the cGAS core domain according to MD simulations. We hypothesize that this tethering is functionally important for realignment of DNA-binding regions of NTD for their

better positioning with core DNA-binding sites, but this statement needs further confirmation. The attempt to decipher the role of α3-helix by titration of NTD with core was unsuccessful, as well as purification of a full-length cGAS protein.

Importantly, while we originally hypothesized sequence specificity that would manifest as distinct chemical shift signatures for different DNA sequences, we observed no sequence-dependent response, but different kinetic regime instead: DNA of comparable length induced fast dynamics, intermediate dynamics, or mixture of both regimes depending on their GC content, with the linear dependency of the amount of exchange broadening from GC content. This argues that NTD can 'sense' physical properties of nuclei acids (such as stiffness, groove geometry, or local structural heterogeneity) rather than sequence itself. This is in line with a recent work by Wang et al. who showed that the activation of cGAS depends on DNA mechanical flexibility, which is controlled by sequence, damage, and length [208]. One plausible interpretation is that increasing GC content increases duplex rigidity, thereby pushing residues into intermediate exchange and broadening peaks without producing qualitatively new chemical environments. A functional role of such an effect is unclear, though increased dwelling time of cGAS on highly rigid DNA molecules might be an additional mechanism of how cGAS distinguished between self- and non-self DNA.

More detailed assessment of the amount and role of intermediate timescale dynamics via CPMG experiments failed due to the suboptimal sample preparation protocol. It is expected that residue-specific $R_{\mathbf{ex}}$ values will follow the same linear dependency as observed in primary analysis based on number of exchange-broadened peaks. Further evaluation of residue-specific values $k_{\mathbf{ex}}$ in complexes with various GC content will elucidate whether the increase in DNA rigidity correlates with the longer $k_{\mathbf{on}}/k_{\mathbf{off}}$ time and therefore results in longer dwelling of NTD on rigid DNA.

4.6 Limitations and future research

Several limitations must be mentioned here: first of all, all NMR experiments were performed on isolated murine NTD, which limits interpretation of results. The formation of the α3-helix was shown only in the SSP plot and in MD simulations, while attempts to assess its role in NTD tethering to cGAS core using full-length cGAS construct and NTD titration with cGAS core have not been successful. In addition, quantitative characterization of intermediate-timescale exchange was impeded by sample preparation issues. Future work should therefore prioritize robust CPMG measurements across DNA panels with controlled GC content and mechanical properties, alongside targeted assessment of α3 helix role in NTD-only and full-length cGAS proteins.

Conclusions

The primary objective of this thesis is to expand the methodological and conceptual approaches available for investigating protein structure and dynamics through solution- and solid-state NMR. Particular emphasis is placed on challenging systems characterized by large molecular weight, limited solubility, conformational heterogeneity, and intrinsically disordered regions. Through three distinct yet conceptually related projects, this research examines how isotopic labeling strategies, sample preparation methods, and integrated analytical frameworks can be optimized to enhance spectral information content and facilitate biologically meaningful interpretation of NMR results.

The first project (Chapter 2) investigated whether cell-extract-based selective deuteration could recover resonances that are inaccessible with standard perdeuterated labeling schemes. The results showed that deuterated algal lysate media facilitate near-complete amide protonation in H_2O and maintain favorable linewidths under fast MAS conditions. In microcrystalline tryptophan synthase from *Salmonella*, this method restored a significant proportion of previously undetectable resonances without loss of signal intensity or spectral resolution, thus achieving the primary goal of enhancing backbone observability in solvent-shielded regions of large oligomeric enzymes.

The second project (Chapter 3) sought to establish a robust and quantitative workflow for preparing TrpB samples suitable for multidimensional MAS NMR and for generating reliable backbone resonance assignments by integrating solution and solid-state datasets. Systematic comparisons of sedimented, microcrystalline, and solution samples, as well as benchmarking batch and vapor-diffusion crystallization methods, identified slowly grown microcrystals as the optimal preparation with respect to resolution, stability, and sensitivity. The acquisition of multiple 4D MAS datasets and the development of a reliability framework directly addressed the objective of evaluating assignment feasibility and internal consistency, resulting in partial backbone assignments for catalytically and allosterically relevant regions. Beyond individual methodological advances, this work contributes a unifying perspective on how sample state—solution, sedimented, or crystalline—modulates both spectral quality and chemical shift behavior. The quantitative comparison of linewidths, peak counts, and solution–MAS chemical shift deviations provides an empirical basis for defining tolerances in automated assignment algorithms and for distinguishing genuine structural perturbations from preparation-induced effects.

The third project (Chapter 4) aimed to elucidate the structure and dynamics of the intrinsically disordered N-terminal domain (NTD) of cGAS. The central research question examined how NTD dynamics, DNA binding, and evolutionary origin contribute to cGAS regulation. Through near-complete backbone assignments, relaxation measurements, titration experiments, and molecular dynamics simulations,

the study demonstrated that the NTD is modular, dynamically heterogeneous, and capable of sensing the physical properties of DNA rather than its sequence. Conceptually, the cGAS NTD study extends the role of NMR beyond static structural description into the realm of evolutionary and dynamic interpretation. The identification of a potential viral origin of NTD and the demonstration that DNA-induced responses depend on mechanical properties rather than sequence challenge existing models of innate immune sensing and suggest new avenues for understanding host–pathogen coevolution.

Future research could further optimize the algal extract-based labeling approach, incorporating controlled concentration series across diverse protein targets. For TrpB, expanding backbone assignments through additional high-dimensional experiments will facilitate a more comprehensive analysis of allosteric networks and catalytic mechanisms. Refining cross-linking strategies by employing milder agents may produce crystals that maintain stability while minimizing spectral perturbation. Regarding cGAS, subsequent studies will leverage quantitative CPMG-based characterization of intermediate-timescale dynamics across DNA panels with defined mechanical properties. Determining how NTD dynamics are coupled to core-domain function remains a critical unresolved question with significant implications for innate immune signaling and viral immune evasion.

Conferences, workshops, and meetings

Conferences

1. ICBS Symposium 2020, 11–13 November 2020, online.
2. 42nd FGMR Annual Discussion Meeting, 27–30 September 2021, online.
3. IMPRS Student Retreat 2022, International Max Planck Research School for Living Matter, 30 March – 2 April 2022, Jugendburg Gemen, Germany.
4. XIIth International conference *NMR: a tool for biology*, 2–4 May 2022, Paris, France.

   Poster presentation: "Allosteric communication in tryptophan synthase studied by ssNMR", H. Kavaleuskaya, S.K. Vasa, R. Linser.
5. European Conference on Magnetic Resonance, 10–14 July 2022, Utrecht, the Netherlands.

   Poster presentation: "Allosteric communication in tryptophan synthase studied by ssNMR", H. Kavaleuskaya, S.K. Vasa, R. Linser.
6. 16th Tag der Chemie 2023, 31 March 2023, TU Dortmund.

   Promoted talk: "NMR study of archaeal protein TrpB", H. Kavaleuskaya, S.K. Vasa, R. Linser.
7. The 2023 Alpine Conference on Magnetic Resonance in Solids, 10–14 September 2023, Chamonix-Mont-Blanc, France.
8. 44th FGMR Annual Discussion Meeting 2023, 18–21 September 2023, Konstanz, Germany.

   Poster presentation: "Chemical shift assignment of 2x43 kDa TrpB from *Pyrococcus furiosus*", H. Kavaleuskaya, S.K. Vasa, R. Linser.
9. IMPRS Student Retreat 2024, International Max Planck Research School for Living Matter, 17–19 April 2024, Meinerzhagen, Germany.

   Promoted talk: "Viral origin of immunity: exploring the disordered domain of cGAS", H. Kavaleuskaya, D. Blokhin, R. Linser.

10. 20$^{\text{th}}$ European Conference on Magnetic Resonance, 30 June–4 July 2024, Bilbao, Spain.

    Poster presentation: “Viral origin of immunity: exploring the disordered domain of cGAS”, H. Kavaleuskaya, D. Blokhin, S.K. Vasa, R. Linser.

11. Annual *bio*-NMR Network Meeting 2024, 27th September 2024, Essen, Germany. Promoted talk: ”Viral origin of immunity: exploring the disordered domain of cGAS”, H. Kavaleuskaya, D. Blokhin, S.K. Vasa, R. Linser.

Workshops

1. Advanced Protein Labeling, 19–22 October 2020, MPI for Molecular Physiology, Dortmund, Germany.

2. 3D Structure Determination and Interpretation: EM/x-ray Crystallography, 8–12 March 2021, MPI for Molecular Physiology, Dortmund, Germany.

3. Good Scientific Practice – Avoid Scientific Misconduct, GSP | ARM!d, M. Mende, 21–23 January 2021, MPI for Molecular Physiology, Dortmund, Germany.

4. Vector Graphics for Scientists (Adobe Illustrator), 7–9 June 2021, MPI for Molecular Physiology, Dortmund, Germany.

Thesis Advisory Committee (TAC) meetings

1. First TAC meeting, 8 June 2021, “Characterization of structure and dynamics in high-molecular-weight drug targets via NMR spectroscopy”.

2. Second TAC meeting, 27 July 2022, “Characterization of structure and dynamics in high-molecular-weight drug targets via NMR spectroscopy”.

3. Third TAC meeting, 9 November 2023, “Backbone resonance assignments and dynamics of thermophilic tryptophan synthase β subunit studied via NMR spectroscopy”.

Bibliography

[1] F. Bloch and A. Siegert, "Magnetic resonance for nonrotating fields," tech. rep., 1940.

[2] F. Bloch, W. W. Hansen, and M. Packard, "The nuclear induction experiment," *Physical Review*, vol. 70, p. 474, 10 1946.

[3] K. Henzler-Wildman and D. Kern, "Dynamic personalities of proteins," 12 2007.

[4] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. analysis of experimental results," *Journal of the American Chemical Society*, vol. 104, no. 17, pp. 4559–4570, 1982.

[5] A. A. Kawale and B. M. Burmann, "Characterization of backbone dynamics using solution nmr spectroscopy to discern the functional plasticity of structurally analogous proteins," *STAR Protocols*, vol. 2, no. 4, p. 100919, 2021.

[6] L. E. Kay, "Nmr studies of protein structure and dynamics," *Journal of Magnetic Resonance*, vol. 173, no. 2, pp. 193–207, 2005.

[7] S. C. L. Kamerlin and A. Warshel, "At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?," *Proteins: Structure, Function, and Bioinformatics*, vol. 78, no. 6, pp. 1339–1375, 2010.

[8] S. D. Schwartz, "Protein dynamics and enzymatic catalysis," *The Journal of Physical Chemistry B*, vol. 127, no. 12, pp. 2649–2660, 2023. PMID: 36944023.

[9] P. Ojeda-May, A. U. Mushtaq, P. Rogne, A. Verma, V. Ovchinnikov, C. Grundström, B. Dulko-Smith, U. H. Sauer, M. Wolf-Watz, and K. Nam, "Dynamic connection between enzymatic catalysis and collective protein motions," *Biochemistry*, vol. 60, no. 28, pp. 2246–2258, 2021. PMID: 34250801.

[10] V. V. Singh Aayushi, Purslow Jeffrey A., "15n cpmg relaxation dispersion for the investigation of protein conformational dynamics on the 181;s-ms timescale," *JoVE*, no. 170, p. e62395, 2021.

[11] S. M. Gobeil, M. C. Ebert, J. Park, D. Gagné, N. Doucet, A. M. Berghuis, J. Pleiss, and J. N. Pelletier, "The structural dynamics of engineered -lactamases vary broadly on three timescales yet sustain native function," *Scientific Reports*, vol. 9, 12 2019.

[12] Y. E. Ryabov and D. Fushman, "A model of interdomain mobility in a multidomain protein," *Journal of the American Chemical Society*, vol. 129, no. 11, pp. 3315–3327, 2007. PMID: 17319663.

[13] P. J. Farber and A. Mittermaier, "Relaxation dispersion nmr spectroscopy for the study of protein allostery," *Biophysical Reviews*, vol. 7, pp. 191–200, 6 2015.

[14] S. Grutsch, S. Brüschweiler, and M. Tollinger, "Nmr methods to study dynamic allostery," *PLoS Computational Biology*, vol. 12, 3 2016.

[15] H. Y. Carr and E. M. Purcell, "Effects of diffusion on free precession in nuclear magnetic resonance experiments," *Phys. Rev.*, vol. 94, pp. 630–638, May 1954.

[16] S. Meiboom and D. Gill, "Modified spin-echo method for measuring nuclear relaxation times," *Review of Scientific Instruments*, vol. 29, pp. 688–691, 08 1958.

[17] M. Akke and A. G. Palmer, "Monitoring macromolecular motions on microsecond to millisecond time scales by r1 −r1 constant relaxation time nmr spectroscopy," *Journal of the American Chemical Society*, vol. 118, no. 4, pp. 911–912, 1996.

[18] M. Dreydoppel, R. J. Lichtenecker, M. Akke, and U. Weininger, "1h r1 relaxation dispersion experiments in aromatic side chains," *Journal of Biomolecular NMR*, vol. 75, pp. 383–392, 12 2021.

[19] A. G. Palmer, C. D. Kroenke, and J. Patrick Loria, "[10] - nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules," in *Nuclear Magnetic Resonance of Biological Macromolecules - Part B* (T. L. James, V. Dötsch, and U. Schmitz, eds.), vol. 339 of *Methods in Enzymology*, pp. 204–238, Academic Press, 2001.

[20] A. Rangadurai, E. S. Szymaski, I. J. Kimsey, H. Shi, and H. M. Al-Hashimi, "Characterizing micro-to-millisecond chemical exchange in nucleic acids using off-resonance r1 relaxation dispersion," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 112-113, pp. 55–102, 2019.

[21] D. F. Hansen, P. Vallurupalli, and L. E. Kay, "An improved 15n relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins," *The Journal of Physical Chemistry B*, vol. 112, no. 19, pp. 5898–5904, 2008. PMID: 18001083.

[22] J. P. Loria, M. Rance, and A. G. Palmer, "A relaxation-compensated carr−purcell−meiboom−gill sequence for characterizing chemical exchange by nmr spectroscopy," *Journal of the American Chemical Society*, vol. 121, no. 10, pp. 2331–2332, 1999.

[23] F. Massi and J. Peng, *Characterizing Protein Dynamics with NMR R 1 Relaxation Experiments*, vol. 1688, pp. 205–221. 01 2018.

[24] T. Orand, E. Delaforge, A. Lee, J. Kragelj, M. Tengo, L. Tengo, M. Blackledge, E. B. Erba, R. J. Davis, A. Palencia, and M. R. Jensen, "Bipartite binding of the intrinsically disordered scaffold protein jip1 to the kinase jnk1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 122, 3 2025.

[25] S. Milles, N. Salvi, M. Blackledge, and M. R. Jensen, "Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments," 12 2018.

[26] T. R. Alderson, E. Adriaenssens, B. Asselbergh, I. Pritišanac, J. V. Lent, H. Y. Gastall, M. A. Wälti, J. M. Louis, V. Timmerman, A. J. Baldwin, and J. L. Benesch, "A weakened interface in the p182l variant of hsp27 associated with severe charcot-marie-tooth neuropathy causes aberrant binding to interacting proteins," *The EMBO Journal*, vol. 40, 4 2021.

[27] M. Ubbink, A. Perrakis, F.-A. Chao, and R. A. Byrd, "Protein dynamics revealed by nmr relaxation methods," *Emerging Topics in Life Sciences*, vol. 2, pp. 93–105, 04 2018.

[28]

[29] J. S. Waugh, "Nmr spectroscopy in solids. a historical perspective," *Analytical Chemistry*, vol. 65, p. 9, 9 1993.

[30] J. H. V. Vleck, "The dipolar broadening of magnetic resonance lines in crystals," *Physical Review*, vol. 74, p. 1168, 11 1948.

[31] E. R. Andrew and R. A. Newing, "The narrowing of nuclear magnetic resonance spectra by molecular rotation in solids," *Proceedings of the Physical Society*, vol. 72, p. 959, dec 1958.

[32] M. Cohen, R. P. Feynman, and L. J. Lowe, "Physical review letters free induction decays of rotating solids," tech. rep., 1957.

[33] P. H. Ysi and C. Review, "Free-induction decays in solids*f i," tech. rep., 1957.

[34] S. Clough and K. W. Gray, "The stochastic theory of the nuclear magnetic resonance line in rotating solids," *Proceedings of the Physical Society*, vol. 79, p. 457, mar 1962.

[35] E. R. Andrew and L. F. Farnell, "The effect of macroscopic rotation on anistotropic bilinear spin interactions in solids," *Molecular Physics*, vol. 15, pp. 157–165, 1968.

[36] J. Schaefer, E. . Stejskal, and R. Buchdahl, "Megic-angle 13cnmr analysis of motion in solid polycarbonate," *Journal of Macromolecular Science, Part B*, vol. 13, pp. 665–672, 5 1977.

[37] J. Schaefer, E. . Stejskal, C. F. Brewer, H. D. Keiser:, and H. Sternlicht', "Cross-polarization 13c nuclear magnetic resonance spectroscopy of collagen," tech. rep., 1978.

[38] E. T. Olejniczak, S. Vega, and R. G. Griffin, "Multiple pulse nmr in rotating solids," *The Journal of Chemical Physics*, vol. 81, pp. 4804–4817, 1984.

[39] A. A. Bax, B. L. Hawkins, G. E. Maciel, and N. M. Szeverenyi, "Two-dimensional nmr spectroscopy," tech. rep., 1984.

[40] F. Creuzet, A. McDermott, R. Gebhard, K. van der Hoef, M. B. Spijker-Assink, J. Herzfeld, J. Lugtenburg, M. H. Levitt, and R. G. Griffin, "Determination of membrane protein structure by rotational resonance nmr: Bacteriorhodopsin," *Science*, vol. 251, pp. 783–786, 2 1991.

[41] P. R. Costa, D. A. Kocisko, B. Q. Sun, . P. T. Lansbury, and R. G. Griffin, "Determination of peptide amide configuration in a model amyloid fibril by solid-state nmr," tech. rep., 1997.

[42] S. Macholl, I. Sack, H. H. Limbach, J. Pauli, M. Kelly, and G. Buntkowsky, "Solid-state nmr study of the sh3 domain of -spectrin: Application of 13c-15n tedor and redor," *Magnetic Resonance in Chemistry*, vol. 38, pp. 596–603, 2000.

[43] L. E. Kay, "The evolution of solution state nmr pulse sequences through the 'eyes' of triple-resonance spectroscopy," *Journal of Magnetic Resonance*, vol. 306, pp. 48–54, 9 2019.

[44] A. T. Petkova, M. Baldus, M. Belenky, M. Hong, R. G. Griffin, and J. Herzfeld, "Backbone and side chain assignment strategies for multiply labeled membrane peptides and proteins in the solid state," *Journal of Magnetic Resonance*, vol. 160, pp. 1–12, 1 2003.

[45] A. Detken, E. H. Hardy, M. Ernst, M. Kainosho, T. Kawakami, S. Aimoto, and B. H. Meier, "Methods for sequential resonance assignment in solid, uniformly 13c, 15n labelled peptides: quantification and application to antamanide," *Journal of biomolecular NMR*, vol. 20, pp. 203–221, 2001.

[46] S. K. Straus, T. Bremi, and R. R. Ernst, "Experiments and strategies for the assignment of fully 13c/15n-labelled polypeptides by solid state nmr," *Journal of biomolecular NMR*, vol. 12, pp. 39–50, 1998.

[47] B. Brutscher, R. Brüschweiler, and R. R. Ernst, "Backbone dynamics and structural characterization of the partially folded a state of ubiquitin by 1h, 13c, and 15n nuclear magnetic resonance spectroscopy," *Biochemistry*, vol. 36, pp. 13043–13053, 10 1997.

[48] A. W. Overhauser, "Polarization of nuclei in metals," *Physical Review*, vol. 92, p. 411, 10 1953.

[49] T. R. Carver and C. P. Slichter, "Experimental verification of the overhauser nuclear polarization effect," *Physical Review*, vol. 102, p. 975, 5 1956.

[50] M. L. Mak-Jurkauskas and R. G. Griffin, "High-frequency dynamic nuclear polarization," *eMagRes*, vol. 2010, 2010.

[51] H. W. Spiess, "Nmr spectroscopy: Pushing the limits of sensitivity," 2008.

[52] R. Verel and B. H. Meier, "Polarization-transfer methods in solid-state magic-angle-spinning nmr: Adiabatic cn pulse sequences," *ChemPhysChem*, vol. 5, pp. 851–862, 6 2004.

[53] M. H. Levitt, "Symmetry in the design of nmr multiple-pulse sequences," *Journal of Chemical Physics*, vol. 128, 2008.

[54] J. Blahut, M. J. Brandl, R. Sarkar, B. Reif, and Z. Tošner, "Optimal control derived sensitivity-enhanced ca-co mixing sequences for mas solid-state nmr – applications in sequential protein backbone assignments," *Journal of Magnetic Resonance Open*, p. 100122, 2023.

[55] A. Bryson, "Applied optimal control: Optimization, estimation and control - a. e. bryson - google books," 5 2018.

[56] M. Bak, J. T. Rasmussen, and N. C. Nielsen, "Simpson: A general simulation program for solid-state nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 147, pp. 296–330, 2000.

[57] H. J. Hogben, M. Krzystyniak, G. T. Charnock, P. J. Hore, and I. Kuprov, "Spinach - a software library for simulation of spin dynamics in large spin systems," *Journal of Magnetic Resonance*, vol. 208, pp. 179–194, 2 2011.

[58] D. Joseph and C. Griesinger, "Optimal control pulses for the 1.2-ghz (28.2-t) nmr spectrometers," tech. rep., 2023.

[59] C. Kehlet, M. Bjerring, A. C. Sivertsen, T. Kristensen, J. J. Enghild, S. J. Glaser, N. Khaneja, and N. C. Nielsen, "Optimal control based nco and nca experiments for spectral assignment in biological solid-state nmr spectroscopy," *Journal of Magnetic Resonance*, vol. 188, pp. 216–230, 2007.

[60] Z. Tošner, R. Sarkar, J. Becker-Baldus, C. Glaubitz, S. Wegner, F. Engelke, S. J. Glaser, and B. Reif, "Overcoming volume selectivity of dipolar recoupling in biological solid-state nmr spectroscopy," *Angewandte Chemie*, vol. 130, pp. 14722–14726, 10 2018.

[61] S. R. Hartmann and E. L. Hahn, "Nuclear double resonance in the rotating frame," *Phys. Rev.*, vol. 128, pp. 2042–2053, Dec 1962.

[62] B. Elena, A. Lesage, S. Steuernagel, A. Böckmann, and L. Emsley, "Proton to carbon-13 inept in solid-state nmr spectroscopy.," *Journal of the American Chemical Society*, vol. 127, pp. 17296–302, 12 2005.

[63] B. Alonso and D. Massiot, "Multi-scale nmr characterisation of mesostructured materials using 1h → 13c through-bond polarisation transfer, fast mas, and 1h spin diffusion," *Journal of Magnetic Resonance*, vol. 163, pp. 347–352, 8 2003.

[64] K. Aebischer and M. Ernst, "Inept and cp transfer efficiencies of dynamic systems in mas solid-state nmr," *Journal of Magnetic Resonance*, vol. 359, p. 107617, 2 2024.

[65] T. Gullion and J. Schaefer, "Rotational-echo double-resonance nmr," *Journal of Magnetic Resonance (1969)*, vol. 81, no. 1, pp. 196–200, 1989.

[66] M. Mehta, D. Gregory, S. Kiihne, D. Mitchell, M. Hatcher, J. Shiels, and G. Drobny, "Distance measurements in nucleic acids using windowless dipolar recoupling solid state nmr," *Solid State Nuclear Magnetic Resonance*, vol. 7, no. 3, pp. 211–228, 1996.

[67] K. E. I. Alla, M A and E. T. Lippmaa, "Selective determination of anisotropic magnetic interactions from high-resolution nmr spectra of powder samples," *Pis'ma Zh. Eksp. Teor. Fiz*, vol. 27:4.

[68] Y. Yarim-Agaev, P. Tutunjian, and J. Waugh, "Sample spinning at the magic angle with rotation-synchronized rf pulses," *Journal of Magnetic Resonance (1969)*, vol. 47, no. 1, pp. 51–60, 1982.

[69] I. Scholz, P. Hodgkinson, B. H. Meier, and M. Ernst, "Understanding two-pulse phase-modulated decoupling in solid-state nmr," *The Journal of Chemical Physics*, vol. 130, p. 114510, 03 2009.

[70] B. Fung, A. Khitrin, and K. Ermolaev, "An improved broadband decoupling sequence for liquid crystals and solids," *Journal of Magnetic Resonance*, vol. 142, no. 1, pp. 97–101, 2000.

[71] J. R. Lewandowski, J. N. Dumez, Ümit Akbey, S. Lange, L. Emsley, and H. Oschkinat, "Enhanced resolution and coherence lifetimes in the solid-state nmr spectroscopy of perdeuterated proteins under ultrafast magic-angle spinning," *Journal of Physical Chemistry Letters*, vol. 2, pp. 2205–2211, 9 2011.

[72] K. H. Gardner and L. E. Kay, "The use of 2h, 13c, 15n multidimensional nmr to study the structure and dynamics of proteins," 1998.

[73] V. Kocman, G. M. D. Mauro, G. Veglia, and A. Ramamoorthy, "Use of paramagnetic systems to speed-up nmr data acquisition and for structural and dynamic studies," 10 2019.

[74] S. Cai, C. Seu, Z. Kovacs, A. D. Sherry, and Y. Chen, "Sensitivity enhancement of multidimensional nmr experiments by paramagnetic relaxation effects," *Journal of the American Chemical Society*, vol. 128, pp. 13474–13478, 10 2006.

[75] K. H. Mroue, R. Zhang, P. Zhu, E. McNerny, D. H. Kohn, M. D. Morris, and A. Ramamoorthy, "Acceleration of natural-abundance solid-state mas nmr measurements on bone by paramagnetic relaxation from gadolinium-dtpa," *Journal of Magnetic Resonance*, vol. 244, pp. 90–97, 2014.

[76] F. Peters, M. Maestre-Martinez, A. Leonov, L. Kovačič, S. Becker, R. Boelens, and C. Griesinger, "Cys-ph-taha: A lanthanide binding tag for rdc and pcs enhanced protein nmr," *Journal of Biomolecular NMR*, vol. 51, pp. 329–337, 11 2011.

[77] M. Strickland, C. D. Schwieters, C. Göbl, A. C. Opina, M. P. Strub, R. E. Swenson, O. Vasalatiy, and N. Tjandra, "Characterizing the magnetic susceptibility tensor of lanthanide-containing polymethylated-dota complexes," *Journal of Biomolecular NMR*, vol. 66, pp. 125–139, 10 2016.

[78] S. J. Ullrich, S. Hölper, and C. Glaubitz, "Paramagnetic doping of a 7tm membrane protein in lipid bilayers by gd 3+-complexes for solid-state nmr spectroscopy," *Journal of Biomolecular NMR*, vol. 58, pp. 27–35, 1 2014.

[79] C. Göbl, M. Resch, M. Strickland, C. Hartlmüller, M. Viertler, N. Tjandra, and T. Madl, "Increasing the chemical-shift dispersion of unstructured proteins with a covalent lanthanide shift reagent," *Angewandte Chemie - International Edition*, vol. 55, pp. 14847–14851, 11 2016.

[80] J.-L. Chen, S. na Chen, H. kai Liu, B. bin Pan, Y. Zhao, X. Fu, G. Otting, and X. cheng Su, "Rational design of lanthanide-binding tags to optimize magnetic anisotropy in paramagnetic protein nmr," *Journal of the American Chemical Society*, vol. 147, pp. 9939–9952, 2025.

[81] R. Linser, V. Chevelkov, A. Diehl, and B. Reif, "Sensitivity enhancement using paramagnetic relaxation in mas solid-state nmr of perdeuterated proteins," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 189, pp. 209–216, 12 2007.

[82] D. P. Frueh, "Practical aspects of nmr signal assignment in larger and challenging proteins," 2014.

[83] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, "Talos+: A hybrid method for predicting protein backbone torsion angles from nmr chemical shifts," *Journal of Biomolecular NMR*, vol. 44, pp. 213–223, 2009.

[84] H. Fraga, C.-A. Arnaud, D. F. Gauto, M. Audin, V. Kurauskas, P. Macek, C. Krichel, J.-Y. Guan, J. Boisbouvier, R. Sprangers, C. Breyton, and P. Schanda, "Solid-state nmr h–n–(c)–h and h–n–c–c 3d/4d correlation experiments for resonance assignment of large proteins," *ChemPhysChem*, vol. 18, no. 19, pp. 2697–2703, 2017.

[85] A. Klein, P. Rovó, V. V. Sakhrani, Y. Wang, J. Holmes, V. Liu, P. Skowronek, L. Kukuk, S. K. Vasa, P. Güntert, L. J. Mueller, R. Linser, A. Klein, P. Rovó, P. Skowronek, S. K. Vasa, R. Linser, L. Kukuk, V. V. Sakhrani, and L. J. Mueller, "Residue-specific insights into (2x)72 kda tryptophan synthase obtained from fast-mas 1 h-detected solid-state nmr," *bioRxiv*, p. 2021.05.12.443859, 5 2021.

[86] H. Aucharova, A. Klein, S. M. Gomez, B. Söldner, S. K. Vasa, and R. Linser, "Protein deuteration via algal amino acids to circumvent proton back-exchange for 1h-detected solid-state nmr," *Chemical Communications*, vol. 60, pp. 3083–3086, 3 2024.

[87] F. Napoli, J. Y. Guan, C. A. Arnaud, P. Macek, H. Fraga, C. Breyton, and P. Schanda, "Deuteration of proteins boosted by cell lysates: High-resolution

amide and h magic-angle-spinning (mas) nmr without the reprotonation bottleneck," *Magnetic Resonance*, vol. 5, pp. 33–49, 4 2024.

[88] A. Eletsky, A. Kienhöfer, and K. Pervushin, "Trosy nmr with partially deuterated proteins," *Journal of Biomolecular NMR*, vol. 20, pp. 177–180, 2001.

[89] T. L. Marchand, T. Schubeis, M. Bonaccorsi, P. Paluch, D. Lalli, A. J. Pell, L. B. Andreas, K. Jaudzems, J. Stanek, and G. Pintacuda, "1h-detected biomolecular nmr under fast magic-angle spinning," 5 2022.

[90] B. Reif, "Deuteration for high-resolution detection of protons in protein magic angle spinning (mas) solid-state nmr," *Chemical Reviews*, vol. 122, pp. 10019–10035, 5 2022.

[91] R. Linser, M. Dasari, M. Hiller, V. Higman, U. Fink, J. M. L. D. Amo, S. Markovic, L. Handel, B. Kessler, P. Schmieder, D. Oesterhelt, H. Oschkinat, and B. Reif, "Proton-detected solid-state nmr spectroscopy of fibrillar and membrane proteins," *Angewandte Chemie (International ed. in English)*, vol. 50, pp. 4508–4512, 5 2011.

[92] S. Asami, P. Schmieder, and B. Reif, "High resolution 1h-detected solid-state nmr spectroscopy of protein aliphatic resonances: access to tertiary structure information," *Journal of the American Chemical Society*, vol. 132, pp. 15133–15135, 11 2010.

[93] C. Opitz, E. Ahrné, K. N. Goldie, A. Schmidt, and S. Grzesiek, "Deuterium induces a distinctive escherichia coli proteome that correlates with the reduction in growth rate," *The Journal of Biological Chemistry*, vol. 294, p. 2279, 2 2018.

[94] S. Aggarwal, C. von Wachenfeldt, S. Z. Fisher, and E. Oksanen, "A protocol for production of perdeuterated ompf porin for neutron crystallography," *Protein Expression and Purification*, vol. 188, p. 105954, 12 2021.

[95] J. J. Katz and H. L. Crespi, "Deuterated organisms: cultivation and uses," *Science (New York, N.Y.)*, vol. 151, pp. 1187–1194, 1966.

[96] B. R. Evans and R. Shah, "Development of approaches for deuterium incorporation in plants," *Methods in enzymology*, vol. 565, pp. 213–243, 2015.

[97] V. Kelpšas and C. von Wachenfeldt, "Enhancing protein perdeuteration by experimental evolution of escherichia coli k-12 for rapid growth in deuterium-based media," *Protein Science : A Publication of the Protein Society*, vol. 30, p. 2457, 12 2021.

[98] M. Cai, Y. Huang, R. Yang, R. Craigie, and G. M. Clore, "A simple and robust protocol for high-yield expression of perdeuterated proteins in escherichia coli grown in shaker flasks," *Journal of biomolecular NMR*, vol. 66, p. 85, 10 2016.

[99] A. Hvidt and S. O. Nielsen, "Hydrogen exchange in proteins," *Advances in protein chemistry*, vol. 21, pp. 287–386, 1966.

[100] P. Y. Cheung, J. E. Churchich, and K. S. Lee, "Refolding of thioredoxin reductase assisted by groel and pdi," *Biochemical and biophysical research communications*, vol. 255, pp. 17–22, 2 1999.

[101] S. M. West, A. D. Guise, and J. B. Chaudhuri, "A comparison of the denaturants urea and guanidine hydrochloride on protein refolding," *Food and Bioproducts Processing*, vol. 75, pp. 50–56, 3 1997.

[102] L. Lecoq, M. Schledorn, S. Wang, S. Smith-Penzel, A. A. Malär, M. Callon, M. Nassal, B. H. Meier, and A. Böckmann, "100 khz mas proton-detected nmr spectroscopy of hepatitis b virus capsids," *Frontiers in Molecular Biosciences*, vol. 6, p. 58, 7 2019.

[103] L. B. Andreas, T. L. Marchand, K. Jaudzems, and G. Pintacuda, "High-resolution proton-detected nmr of proteins at very fast mas," *Journal of Magnetic Resonance*, vol. 253, pp. 36–49, 4 2015.

[104] J. Medeiros-Silva, D. Mance, M. Daniëls, S. Jekhmane, K. Houben, M. Baldus, and M. Weingarth, "1h-detected solid-state nmr studies of water-inaccessible proteins in vitro and in situ," *Angewandte Chemie - International Edition*, vol. 55, pp. 13606–13610, 10 2016.

[105] M. Takeda and M. Kainosho, "Cell-free protein production for nmr studies," *Methods in molecular biology (Clifton, N.J.)*, vol. 831, pp. 71–84, 2012.

[106] G. David, M. L. Fogeron, M. Schledorn, R. Montserret, U. Haselmann, S. Penzel, A. Badillo, L. Lecoq, P. André, M. Nassal, R. Bartenschlager, B. H. Meier, and A. Böckmann, "Structural studies of self-assembled subviral particles: Combining cell-free expression with 110 khz mas nmr spectroscopy," *Angewandte Chemie International Edition*, vol. 57, pp. 4787–4791, 4 2018.

[107] F. Löhr, V. Katsemi, J. Hartleib, U. Günther, and H. Rüterjans, "A strategy to obtain backbone resonance assignments of deuterated proteins in the presence of incomplete amide 2 h/ 1 h back-exchange," tech. rep., 2003.

[108] K. T. Movellan, E. E. Najbauer, S. Pratihar, M. Salvi, K. Giller, S. Becker, and L. B. Andreas, "Alpha protons as nmr probes in deuterated proteins," *Journal of Biomolecular NMR*, vol. 73, pp. 81–91, 2 2019.

[109] E. Hilario, L. Fan, L. J. Mueller, and M. F. Dunn, "on," *Journal of Visualized Experiments*, vol. 2020, pp. 1–25, 2020.

[110] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, and G. W. Vuister, "Ccpnmr analysisassign: a flexible platform for integrated nmr analysis," *Journal of Biomolecular NMR*, vol. 66, pp. 111–124, 10 2016.

[111] A. Vallet, A. Favier, B. Brutscher, and P. Schanda, "ssnmrlib: a comprehensive library and tool box for acquisition of solid-state nuclear magnetic resonance experiments on bruker spectrometers," *Magnetic Resonance*, vol. 1, pp. 331–345, 2020.

[112] D. Bell, F. Lindemann, L. Gerland, H. Aucharova, A. Klein, D. Friedrich, M. Hiller, K. Grohe, B. van Rossum, A. Diehl, J. Hughes, L. J. Mueller, R. Linser, A.-F. Miller, and H. Oschkinat, "Sedimentation of large, soluble proteins up to 140 kda for 1h-detected mas nmr and 13c dnp nmr – practical aspects," 2 2024.

[113] A. Marchetti, S. Jehle, M. Felletti, M. J. Knight, Y. Wang, Z. Q. Xu, A. Y. Park, G. Otting, A. Lesage, L. Emsley, N. E. Dixon, and G. Pintacuda, "Backbone assignment of fully protonated solid proteins by 1h detection and ultrafast magic-angle-spinning nmr spectroscopy," *Angewandte Chemie International Edition*, vol. 51, pp. 10756–10759, 10 2012.

[114] M. J. Knight, A. L. Webber, A. J. Pell, P. Guerry, E. Barbet-Massin, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, L. Emsley, A. Lesage, T. Herrmann, and G. Pintacuda, "Fast resonance assignment and fold determination of human superoxide dismutase by high-resolution proton-detected solid-state mas nmr spectroscopy," *Angewandte Chemie - International Edition*, vol. 50, pp. 11697–11701, 12 2011.

[115] M. Benvenuti and S. Mangani, "Crystallization of soluble proteins in vapor diffusion for x-ray crystallography," *Nature Protocols 2007 2:7*, vol. 2, pp. 1633–1651, 6 2007.

[116] J. D. Ng, B. Lorber, J. Witz, A. Théobald-Dietrich, D. Kern, and R. Giegé, "The crystallization of biological macromolecules from precipitates: evidence for ostwald ripening," *Journal of Crystal Growth*, vol. 168, pp. 50–62, 10 1996.

[117] T. Gopinath, S. E. Nelson, and G. Veglia, "1h-detected mas solid-state nmr experiments enable the simultaneous mapping of rigid and dynamic domains of membrane proteins," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 285, pp. 101–107, 12 2017.

[118] I. Bertini, C. Luchinat, G. Parigi, E. Ravera, B. Reif, and P. Turano, "Solid-state nmr of proteins sedimented by ultracentrifugation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 10396–10399, 6 2011.

[119] I. Bertini, C. Luchinat, G. Parigi, and E. Ravera, "Sednmr: On the edge between solution and solid-state nmr," *Accounts of Chemical Research*, vol. 46, pp. 2059–2069, 9 2013.

[120] Z. Liu, W. P. Zhang, Q. Xing, X. Ren, M. Liu, and C. Tang, "Noncovalent dimerization of ubiquitin," *Angewandte Chemie (International ed. in English)*, vol. 51, pp. 469–472, 1 2012.

[121] F. Shi, T. Zhang, J. Li, C. Shi, and S. Xiang, "Studying large biomolecules as sedimented solutes with solid-state nmr," *Biophysics Reports*, vol. 10, p. 201, 8 2024.

[122] M. Fragai, C. Luchinat, G. Parigi, and E. Ravera, "Practical considerations over spectral quality in solid state nmr spectroscopy of soluble proteins," *Journal of Biomolecular NMR*, vol. 57, pp. 155–166, 10 2013.

[123] S. M. Gomez, *Characterization of protein structure and dynamics by solution and solid-state NMR*. PhD thesis, Technische Universität Dortmund, 2023.

[124] I. R. Krauss, G. Ferraro, A. Pica, J. A. Márquez, J. R. Helliwell, and A. Merlino, "Principles and methods used to grow and optimize crystals of protein–metallodrug adducts, to determine metal binding sites and to assign metal ligands," *Metallomics*, vol. 9, pp. 1534–1547, 11 2017.

[125] D. W. Pack, G. Chen, K. M. Maloney, C. T. Chen, and F. H. Arnold, "A metal-chelating lipid for 2d protein crystallization via coordination of surface histidines," *Journal of the American Chemical Society*, vol. 119, pp. 2479–2487, 3 1997.

[126] T. E. Edwards, R. Liao, I. Phan, P. J. Myler, and C. Grundner, "Mycobacterium thermoresistibile as a source of thermostable orthologs of mycobacterium tuberculosis proteins," *Protein Science*, vol. 21, pp. 1093–1096, 7 2012.

[127] A. M. Facchiano, G. Colonna, and R. Ragone, "Helix stabilizing factors and stabilization of thermophilic proteins: an x-ray based study," *Protein engineering*, vol. 11, pp. 753–760, 1998.

[128] Y. X. Fan, P. McPhie, and E. W. Miles, "Regulation of tryptophan synthase by temperature, monovalent cations, and an allosteric ligand. evidence from arrhenius plots, absorption spectra, and primary kinetic isotope effects," *Biochemistry*, vol. 39, pp. 4692–4703, 4 2000.

[129] E. Watkins-Dulaney, S. Straathof, and F. Arnold, "Tryptophan synthase: Biocatalyst extraordinaire," 2020.

[130] A. R. Buller, P. V. Roye, J. K. Cahn, R. A. Scheele, M. Herger, and F. H. Arnold, "Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble," *Journal of the American Chemical Society*, vol. 140, pp. 7256–7266, 6 2018.

[131] A. R. Buller, S. Brinkmann-Chen, D. K. Romney, M. Herger, J. Murciano-Calles, and F. H. Arnold, "Directed evolution of the tryptophan synthase -subunit for stand-alone function recapitulates allosteric activation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, 2015.

[132] J. Stanek and W. Koźmiński, "Iterative algorithm of discrete fourier transform for processing randomly sampled nmr data sets," *Journal of Biomolecular NMR*, vol. 47, pp. 65–77, 5 2010.

[133] B. Han, Y. Liu, S. W. Ginzinger, and D. S. Wishart, "Shiftx2: significantly improved protein chemical shift prediction," *Journal of Biomolecular Nmr*, vol. 50, p. 43, 5 2011.

[134] E. Schmidt and P. Güntert, "A new algorithm for reliable and general nmr resonance assignment," *Journal of the American Chemical Society*, vol. 134, pp. 12817–12829, 8 2012.

[135] C. J. Lusty, "A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography," *urn:issn:0021-8898*, vol. 32, pp. 106–112, 2 1999.

[136] A. Klein, *Development and application of new 1H-detected solid-state NMR methods to expand dimensionality and study protein-ligand interactions*. PhD thesis, Technische Universität Dortmund, 2021.

[137] H. Aucharova and R. Linser, "Assignment of the n-terminal domain of mouse cgas," *Biomolecular NMR Assignments*, vol. 19, pp. 35–39, 6 2025.

[138] M. Li and H. B. Shu, "Dephosphorylation of cgas by ppp6c impairs its substrate binding activity and innate antiviral response," *Protein and Cell*, vol. 11, 2020.

[139] Z. Cheng, T. Dai, X. He, Z. Zhang, F. Xie, S. Wang, L. Zhang, and F. Zhou, "The interactions between cgas-sting pathway and pathogens," 2020.

[140] A. P. West and G. S. Shadel, "Mitochondrial dna in innate immune responses and inflammatory pathology," 6 2017.

[141] P. J. Kranzusch, "cgas and cd-ntase enzymes: structure, mechanism, and evolution," 12 2019.

[142] M. Gentili, X. Lahaye, F. Nadalin, G. F. Nader, E. P. Lombardi, S. Herve, N. S. D. Silva, D. C. Rookhuizen, E. Zueva, C. Goudot, M. Maurin, A. Bochnakian, S. Amigorena, M. Piel, D. Fachinetti, A. Londoño-Vallejo, and N. Manel, "The n-terminal domain of cgas determines preferential association with centromeric dna and innate immune activation in the nucleus," *Cell Reports*, vol. 26, pp. 2377–2393.e13, 2 2019.

[143] H. Yang, H. Wang, U. Ren, Q. Chen, and Z. J. Chena, "Cgas is essential for cellular senescence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, pp. E4612–E4620, 6 2017.

[144] J. A. Boyer, C. J. Spangler, J. D. Strauss, A. P. Cesmat, P. Liu, R. K. McGinty, and Q. Zhang, "Structural basis of nucleosome-dependent cgas inhibition," *Science*, 2020.

[145] T. Li, T. Huang, M. Du, X. Chen, F. Du, J. Ren, and Z. J. Chen, "Phosphorylation and chromatin tethering prevent cgas activation during mitosis," *Science*, vol. 371, 3 2021.

[146] S. Michalski, C. C. de Oliveira Mann, C. A. Stafford, G. Witte, J. Bartho, K. Lammens, V. Hornung, and K. P. Hopfner, "Structural basis for sequestration and autoinhibition of cgas by chromatin," *Nature*, vol. 587, pp. 678–682, 11 2020.

[147] X. Duan, Y. Yang, S. Wang, X. Feng, T. Wang, P. Wang, M. Ding, H. Zhang, B. Liu, W. Wei, W. Yao, L. Cui, X. Zhou, and W. Wang, "Dose-related telomere damage associated with the genetic polymorphisms of cgas/sting signaling pathway in the workers exposed by pahs," *Environmental Pollution*, vol. 260, 5 2020.

[148] H. Liu, H. Zhang, X. Wu, D. Ma, J. Wu, L. Wang, Y. Jiang, Y. Fei, C. Zhu, R. Tan, P. Jungblut, G. Pei, A. Dorhoi, Q. Yan, F. Zhang, R. Zheng, S. Liu, H. Liang, Z. Liu, H. Yang, J. Chen, P. Wang, T. Tang, W. Peng, Z. Hu, Z. Xu, X. Huang, J. Wang, H. Li, Y. Zhou, F. Liu, D. Yan, S. H. Kaufmann, C. Chen, Z. Mao, and B. Ge, "Nuclear cgas suppresses dna repair and promotes tumorigenesis," *Nature*, vol. 563, pp. 131–136, 11 2018.

[149] H. Chen, J. Zhang, Y. Wang, A. Simoneau, H. Yang, A. S. Levine, L. Zou, Z. Chen, and L. Lan, "cgas suppresses genomic instability as a decelerator of replication forks," tech. rep., 2020.

[150] J. Li and R. A. Byrd, "A simple protocol for the production of highly deuterated proteins for biophysical studies," *The Journal of Biological Chemistry*, vol. 298, 8 2022.

[151] J. C. Martinez, F. Morandini, L. Fitzgibbons, N. Sieczkiewicz, S. J. Bae, M. E. Meadow, E. Hillpot, J. Cutting, V. Paige, S. A. Biashad, M. Simon, J. Sedivy, A. Seluanov, and V. Gorbunova, "cgas deficient mice display premature aging associated with de-repression of line1 elements and inflammation," 10 2024.

[152] A. M. Herzner, C. A. Hagmann, M. Goldeck, S. Wolter, K. Kübler, S. Wittmann, T. Gramberg, L. Andreeva, K. P. Hopfner, C. Mertens, T. Zillinger, T. Jin, T. S. Xiao, E. Bartok, C. Coch, D. Ackermann, V. Hornung, J. Ludwig, W. Barchet, G. Hartmann, and M. Schlee, "Sequence-specific activation of the dna sensor cgas by y-form dna structures as found in primary hiv-1 cdna," *Nature Immunology*, vol. 16, pp. 1025–1033, 9 2015.

[153] L. Andreeva, *Cytosolic DNA sensing via cGAS: long, U-shaped and other DNA ligands and cellular co-factors*. PhD thesis, Ludwig-Maximilians-Universität München, 2018.

[154] P. Xia, S. Wang, B. Ye, Y. Du, C. Li, Z. Xiong, Y. Qu, and Z. Fan, "A circular rna protects dormant hematopoietic stem cells from dna sensor cgas-mediated exhaustion," *Immunity*, vol. 48, pp. 688–701.e7, 4 2018.

[155] M. Du and Z. J. Chen, "Dna-induced liquid phase condensation of cgas activates innate immune signaling," *Science*, vol. 361, pp. 704–709, 8 2018.

[156] S. Chen, M. Rong, Y. Lv, D. Zhu, and Y. Xiang, "Regulation of cgas activity by rna-modulated phase separation," *EMBO reports*, vol. 24, 2 2023.

[157] L. Zhong, M. M. Hu, L. J. Bian, Y. Liu, Q. Chen, and H. B. Shu, "Phosphorylation of cgas by cdk1 impairs self-dna sensing in mitosis," *Cell Discovery*, vol. 6, 2020.

[158] Z. M. Song, H. Lin, X. M. Yi, W. Guo, M. M. Hu, and H. B. Shu, "Kat5 acetylates cgas to promote innate immune response to dna virus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 21568–21575, 2020.

[159] J. Tao, X.-W. Zhang, J. Jin, X.-X. Du, T. Lian, J. Yang, X. Zhou, Z. Jiang, and X.-D. Su, "Nonspecific dna binding of cgas n terminus promotes cgas activation," *The Journal of Immunology*, vol. 198, pp. 3627–3636, 2017.

[160] A. Lee, E. B. Park, J. Lee, B. S. Choi, and S. J. Kang, "The n terminus of cgas de-oligomerizes the cgas:dna complex and lifts the dna size restriction of core-cgas activity," *FEBS Letters*, vol. 591, pp. 954–961, 3 2017.

[161] Z. Jiang, F. Shi, J. Li, R. Liu, J. Zhou, Z. Zhong, C. Shi, M. Ma, S. Xiang, and D. Gao, "Crucial role of the cgas n terminus in mediating flowable and functional cgas–dna condensate formation via dna interactions," *Proceedings of the National Academy of Sciences*, vol. 122, 1 2025.

[162] P. J. Kranzusch, A. S. Y. Lee, J. M. Berger, and J. A. Doudna, "Structure of human cgas reveals a conserved family of second-messenger enzymes in innate immunity," *Cell Reports*, vol. 3, pp. 1362–1368, 5 2013.

[163] B. D. Schweibenz, S. C. Devarkar, M. Solotchi, C. Craig, J. Zheng, B. D. Pascal, S. Gokhale, P. Xie, P. R. Griffin, and S. S. Patel, "The intrinsically disordered cards-helicase linker in rig-i is a molecular gate for rna proofreading," *The EMBO Journal*, vol. 41, p. e109782, 5 2022.

[164] E. Sayers, J. Beck, E. Bolton, J. Brister, J. Chan, R. Connor, M. Feldgarden, A. Fine, K. Funk, J. Hoffman, S. Kannan, C. Kelly, W. Klimke, S. Kim, S. Lathrop, A. Marchler-Bauer, T. Murphy, C. O'Sullivan, E. Schmieder, Y. Skripchenko, A. Stine, F. Thibaud-Nissen, J. Wang, J. Ye, E. Zellers, V. Schneider, and K. Pruitt, "Database resources of the national center for biotechnology information in 2025," *Nucleic Acids Research*, vol. 53, pp. D20–D29, 11 2024.

[165] T. U. Consortium, "Uniprot: the universal protein knowledgebase," *Nucleic Acids Research*, vol. 45, pp. D158–D169, 11 2016.

[166] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks (amino add sequence/alignment algorithms/data base srching)," tech. rep., 1992.

[167] S. Kumar, G. Stecher, M. Suleski, M. Sanderford, S. Sharma, and K. Tamura, "Mega12: Molecular evolutionary genetic analysis version 12 for adaptive and green computing," *Molecular Biology and Evolution*, vol. 41, 12 2024.

[168] J. Castresana, "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis," *Mol. Biol. Evol*, vol. 17, pp. 540–552, 2000.

[169] G. Baele, X. Ji, G. W. Hassler, J. T. McCrone, Y. Shao, Z. Zhang, A. J. Holbrook, P. Lemey, A. J. Drummond, A. Rambaut, and M. A. Suchard, "Beast x for bayesian phylogenetic, phylogeographic and phylodynamic inference," *Nature Methods*, vol. 22, pp. 1653–1656, 8 2025.

[170] C. L. Schoch, S. Ciufo, M. Domrachev, C. L. Hotton, S. Kannan, R. Khovanskaya, D. Leipe, R. McVeigh, K. O'Neill, B. Robbertse, S. Sharma, V. Soussov, J. P. Sullivan, L. Sun, S. Turner, and I. Karsch-Mizrachi, "Ncbi taxonomy: A comprehensive update on curation, resources and tools," *Database*, vol. 2020, 2020.

[171] F. Madeira, N. Madhusoodanan, J. Lee, A. Eusebi, A. Niewielska, A. R. N. Tivey, R. Lopez, and S. Butcher, "The embl-ebi job dispatcher sequence analysis tools framework in 2024," *Nucleic acids research*, vol. 52, p. W521—W525, July 2024.

[172] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, and J. M. Jumper, "Accurate structure prediction of biomolecular interactions with alphafold 3," *Nature*, vol. 630, pp. 493–500, 6 2024.

[173] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, "Ucsf chimerax: Structure visualization for researchers, educators, and developers," *Protein Science*, vol. 30, no. 1, pp. 70–82, 2021.

[174] S. Contreras-Martos, H. H. Nguyen, P. N. Nguyen, N. Hristozova, M. Macossay-Castillo, D. Kovacs, A. Bekesi, J. S. Oemig, D. Maes, K. Pauwels, P. Tompa, and P. Lebrun, "Quantification of intrinsically disordered proteins: A problem not fully appreciated," *Frontiers in Molecular Biosciences*, vol. 5, 9 2018.

[175] C. Bracken, A. G. Palmer, and J. Cavanagh, "(h)n(coca)nh and hn(coca)nh experiments for 1h-15n backbone assignments in 13c/15n-labeled proteins," *Journal of Biomolecular NMR*, vol. 9, pp. 94–100, 1997.

[176] A. G. Palmer, "Dynamic properties of proteins from nmr spectroscopy," tech. rep., 1993.

[177] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. V. D. Spoel, B. Hess, and E. Lindahl, "Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, pp. 845–854, 4 2013.

[178] J. Huang and A. D. Mackerell, "Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data," *Journal of Computational Chemistry*, vol. 34, pp. 2135–2145, 9 2013.

[179] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, pp. 926–935, 07 1983.

[180] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, 10 1984.

[181] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, 12 1981.

[182] J. T. Nielsen and F. A. Mulder, "Potenci: Prediction of temperature, neighbor and ph-corrected chemical shifts for intrinsically disordered proteins," *Journal of Biomolecular NMR*, vol. 70, pp. 141–165, 3 2018.

[183] J. T. Nielsen and F. A. Mulder, "Chespi: chemical shift secondary structure population inference," *Journal of biomolecular NMR*, vol. 75, pp. 273–291, 7 2021.

[184] C. C. D. O. Mann, R. Kiefersauer, G. Witte, and K. P. Hopfner, "Structural and biochemical characterization of the cell fate determining nucleotidyltransferase fold protein mab21l1," *Scientific Reports*, vol. 6, pp. 1–14, 6 2016.

[185] A. L. Bruel, A. Masurel-Paulet, J. B. Rivière, Y. Duffourd, D. Lehalle, C. Bensignor, F. Huet, J. Borgnon, F. Roucher, P. Kuentz, J. F. Deleuze, C. Thauvin-Robinet, L. Faivre, and J. Thevenon, "Autosomal recessive truncating mab21l1 mutation associated with a syndromic scrotal agenesis," *Clinical Genetics*, vol. 91, pp. 333–338, 2 2017.

[186] J. Rainger, D. Pehlivan, S. Johansson, H. Bengani, L. Sanchez-Pulido, K. A. Williamson, M. Ture, H. Barker, K. Rosendahl, J. Spranger, D. Horn, A. Meynert, J. A. Floyd, T. Prescott, C. A. Anderson, J. K. Rainger, E. Karaca, C. Gonzaga-Jauregui, S. Jhangiani, D. M. Muzny, A. Seawright, D. C. Soares, M. Kharbanda, V. Murday, A. Finch, R. A. Gibbs, V. V. Heyningen, M. S. Taylor, T. Yakut, P. M. Knappskog, M. E. Hurles, C. P. Ponting, J. R. Lupski, G. Houge, and D. R. Fitzpatrick, "Monoallelic and biallelic mutations in mab21l2 cause a spectrum of major eye malformations," *American Journal of Human Genetics*, vol. 94, pp. 915–923, 6 2014.

[187] B. Deml, A. Kariminejad, R. H. Borujerdi, S. Muheisen, L. M. Reis, and E. V. Semina, "Mutations in mab21l2 result in ocular coloboma, microcornea and cataracts," *PLoS Genetics*, vol. 11, pp. 1–26, 2015.

[188] J. Sridharan, T. Haremaki, Y. Jin, S. Teegala, and D. C. Weinstein, "Xmab21l3 mediates dorsoventral patterning in xenopus laevis," *Mechanisms of Development*, vol. 129, pp. 136–146, 7 2012.

[189] C. Takahashi, M. Kusakabe, T. Suzuki, K. Miyatake, and E. Nishida, "Mab21-l3 regulates cell fate specification of multiciliate cells and ionocytes," *Nature Communications*, vol. 6, 2015.

[190] T. Ogami, Y. Tamura, K. Toss, K. Yuki, M. Morikawa, S. Tsutsumi, H. Aburatani, K. Miyazawa, K. Miyazono, and D. Koinuma, "Mab21l4 regulates the tgf- -induced expression of target genes in epidermal keratinocytes," *Journal of Biochemistry*, vol. 171, pp. 399–410, 4 2022.

[191] H. Liu, Z. Yan, D. Zhu, H. Xu, F. Liu, T. Chen, H. Zhang, Y. Zheng, B. Liu, L. Zhang, W. Zhao, and C. Gao, "Cd-ntase family member mb21d2 promotes

cgas-mediated antiviral and antitumor immunity," *Cell Death and Differentiation*, vol. 30, pp. 992–1004, 4 2023.

[192] S.-H. W. Y.-G. C. C.-L. H. S.-F. H. L.-Y. S. H.-F. Y.-Y. J. J.-Y. Y. Y.-J. C. Wei-Ting Liao, Kun-Chin Ho, "Cbap modulates akt-dependent tsc2 phosphorylation to promote rheb-mtorc1 signaling and growth of t-cell acute lymphoblastic leukemia," *Oncogene*, vol. 38, pp. 1432–1447, 9 2019.

[193] Y. J. Chiang, K. C. Ho, C. T. Sun, J. J. Chiu, F. J. Lee, F. Liao, H. F. Yang-Yen, and J. J. Y. Yen, "Cbap functions as a novel component in chemokine-induced zap70-mediated t-cell adhesion and migration," *PLoS ONE*, vol. 8, 4 2013.

[194] C. J. Kao, Y. J. Chiang, P. H. Chen, K. R. Lin, P. I. Hwang, H. F. Yang-Yen, and J. J. Yen, "Cbap interacts with the un-liganded common -subunit of the gm-csf/il-3/il-5 receptor and induces apoptosis via mitochondrial dysfunction," *Oncogene*, vol. 27, pp. 1397–1403, 2 2008.

[195] V. Richter, C. S. Palmer, L. D. Osellame, A. P. Singh, K. Elgass, D. A. Stroud, H. Sesaki, M. Kvansakul, and M. T. Ryan, "Structural and functional analysis of mid51, a dynamin receptor required for mitochondrial fission," *Journal of Cell Biology*, vol. 204, pp. 477–486, 2014.

[196] O. C. Losón, R. Liu, M. E. Rome, S. Meng, J. T. Kaiser, S. O. Shan, and D. C. Chan, "The mitochondrial fission receptor mid51 requires adp as a cofactor," *Structure*, vol. 22, pp. 367–377, 3 2014.

[197] S. Qiu, X. Zhong, X. Meng, S. Li, X. Qian, H. Lu, J. Cai, Y. Zhang, M. Wang, Z. Ye, H. Zhang, and P. Gao, "Mitochondria-localized cgas suppresses ferroptosis to promote cancer progression," *Cell Research*, vol. 33, pp. 299–311, 4 2023.

[198] K. Mosallanejad, S. N. Kennedy, K. M. Bahleda, K. M. Slavik, W. Zhou, A. A. Govande, D. C. Hancks, P. J. Kranzusch, and J. C. Kagan, "Species-specific self-dna detection mechanisms by mammalian cyclic gmp-amp synthases," *Science Immunology*, vol. 8, 1 2023.

[199] B. C. Campbell, M. G. Paez-Segala, L. L. Looger, G. A. Petsko, and C. F. Liu, "Chemically stable fluorescent proteins for advanced microscopy," *Nature Methods*, vol. 19, pp. 1612–1621, 12 2022.

[200] K. Tamiola, B. Acar, and F. A. Mulder, "Sequence-specific random coil chemical shifts of intrinsically disordered proteins," *Journal of the American Chemical Society*, vol. 132, 2010.

[201] D. Saha, S. Hailu, A. Hada, J. Lee, J. Luo, J. A. Ranish, Y. chi Lin, K. Feola, J. Persinger, A. Jain, B. Liu, Y. Lu, P. Sen, and B. Bartholomew, "The at-hook is an evolutionarily conserved auto-regulatory domain of swi/snf required for cell lineage priming," *Nature Communications*, vol. 14, 12 2023.

[202] N. E. Davey, K. V. Roey, R. J. Weatheritt, G. Toedt, B. Uyar, B. Altenberg, A. Budd, F. Diella, H. Dinkel, and T. J. Gibson, "Attributes of short linear motifs," *Molecular BioSystems*, vol. 8, pp. 268–281, 1 2012.

[203] K. V. Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, and N. E. Davey, "Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation," 7 2014.

[204] D. J. McGeoch, F. J. Rixon, and A. J. Davison, "Topics in herpesvirus genomics and evolution," *Virus Research*, vol. 117, no. 1, pp. 90–104, 2006. Comparative Genomics and Evolution of Complex Viruses.

[205] M. Huang, K. Song, X. Liu, S. Lu, Q. Shen, R. Wang, J. Gao, Y. Hong, Q. Li, D. Ni, J. Xu, G. Chen, and J. Zhang, "Allofinder: A strategy for allosteric modulator discovery and allosterome analyses," *Nucleic Acids Research*, vol. 46, pp. W451–W458, 7 2018.

[206] J.-J. Xu, X.-F. Zhang, Y. Jiang, H. Fan, J.-X. Li, C.-Y. Li, Q. Zhao, L. Yang, Y.-H. Hu, C. Martin, and X.-Y. Chen, "A unique flavoenzyme operates in ubiquinone biosynthesis in photosynthesis-related eukaryotes," tech. rep., 2021.

[207] J. Zhang, J. Zhao, S. Xu, J. Li, S. He, Y. Zeng, L. Xie, N. Xie, T. Liu, K. Lee, G. J. Seo, L. Chen, A. C. Stabell, Z. Xia, S. L. Sawyer, J. Jung, C. Huang, and P. Feng, "Species-specific deamidation of cgas by herpes simplex virus ul37 protein facilitates viral replication," *Cell Host and Microbe*, vol. 24, pp. 234–248.e5, 8 2018.

[208] L. Wang, S. Li, K. Wang, N. Wang, Q. Liu, Z. Sun, L. Wang, L. Wang, Q. Liu, C. Song, C. Liu, and Q. Yang, "Dna mechanical flexibility controls dna potential to activate cgas-mediated immune surveillance," *Nature Communications*, vol. 13, 12 2022.

List of Figures

List of Tables

Appendix A

Supplementary Tables

Table A.1: Summary of quantum mechanical parameters describing spin system.

Quantum number	**Symbol**	**Description**	**Values**	**Physical interpretation**
Mass number	A	Total nucleons (protons + neutrons)	Any positive integer	
Atomic number	Z	Number of protons	Any positive integer ≥ 1	Defines the net charge of the nucleus
Neutron number	N	Number of neutrons	Any non-negative integer	
Magnetic quantum number	m_l	Projection of l along the Z-axis	$-l, -l+1, \ldots, l-1, l$	Describes how much angular momentum is aligned with the external magnetic field; determines how many sublevels exist for each orbital.
Orbital quantum number	L	Orbital angular momentum of a nucleon	$0, 1, 2, 3, \ldots$	If $L = 0 \rightarrow$ the nucleon moves in a spherical orbital. If $L > 0 \rightarrow$ the nucleon has rotational motion around the centre of the nucleus.
magnitude $\rightarrow$	$\mathbf{L}$	Magnitude of orbital quantum number	$\lvert\mathbf{L}\rvert = \hbar\sqrt{L(L+1)}$	Represents the total nuclear spin angular momentum due to the orbital motion of a nucleon around the nucleus. Influences energy level splittings and spin–orbit coupling, impacting relaxation dynamics in NMR.
Nuclear spin number	I	Total angular momentum of the entire nucleus	$0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}, 3, \ldots$ (depends on nucleons and pairing)	The total angular momentum of the entire nucleus, considering all nucleons. Nuclear spin arises from the coupling of all individual nucleons' angular momenta.
magnitude $\rightarrow$	$\mathbf{I}$	Magnitude of total angular momentum	$\lvert\mathbf{I}\rvert = \hbar\sqrt{I(I+1)}$	Represents the absolute value of the total angular momentum vector. Affects the energy level splitting in a magnetic field and influences the transition probabilities.
Spin	S	Intrinsic spin of a nucleon	Always $S = \frac{1}{2}$ for fermions (protons, neutrons)	Intrinsic property of elementary particles, fundamental source of angular momentum; determines how particles behave under magnetic fields.
Total angular momentum	J	$J = L + S$	$J = L - S, (L-S)+1, \ldots, L+S$	Combination of orbital angular momentum (L) and intrinsic spin (S); describes how an individual proton or neutron moves inside the nucleus.

Table A.2: Properties of Hamiltonians of different NMR interactions.

Interaction	**Description**	**Mathematical form**	**Motion dependence**	**Effect on NMR**
Zeeman interaction	Describes nuclear spin interaction with an external magnetic field.	$\hat{\mathcal{H}}_Z = \text{-}\gamma \text{B}_0 \hat{I}_z$	No (Zeeman interaction is static).	Always present in all NMR experiments.
$(\pi/2)_x$ pulse	Describes spin evolution under a radiofrequency (RF) pulse along the x-axis.	$\hat{\mathcal{H}}_x = \text{-}\gamma \text{B}_1 \hat{I}_x$	No (pulse-driven evolution).	Used in pulse sequences to manipulate magnetization.
J-coupling	Describes indirect through-bond spin–spin interaction.	$\hat{\mathcal{H}}_Z = 2\pi \text{J}\ _z \hat{S}_z$	No (through-bond interaction is rigid).	Common in solution NMR (HSQC, COSY, TOCSY).
Dipolar interaction	Describes through-space dipolar interactions between two spins.	$\hat{\mathcal{H}}_D{}^{\text{secular}} = -\frac{\mu_0}{4\pi}\frac{\hbar\gamma_1\gamma_2}{r^3}\left[\frac{3\cos^2\theta - 1}{2}\right]\hat{I}_z\hat{S}_z$	Yes (depends on molecular tumbling).	Important in solid-state NMR and relaxation studies.
RDC	Describes dipolar couplings that are partially averaged in partially aligned media.	$\hat{\mathcal{H}}_{\text{RDC}} = D(3\cos^2\theta - 1)\hat{I}_z\hat{S}_z$	Yes (affected by molecular alignment).	Used in alignment media for structural studies.
NOE	Describes cross-relaxation and spin polarization transfer in NOE experiments.	$\hat{\mathcal{H}}_{\text{NOE}} = \omega_I \hat{I}_z + \omega_S \hat{S}_z + \sigma(\hat{I}_z\hat{S}_z)$	Yes (strongly depends on molecular motion).	Used in NOESY, ROESY, and biomolecular NMR experiments.

Table A.3: Comparison of size of interactions having effect on solution- and solid-state NMR spectra.

Parameter	**Solution NMR (@ 800 MHz, 5 mm NMR tube, ^{2}H,^{13}C,^{15}N-sample)**	**ssNMR (@ 800 MHz, 55.5 kHz spinning, 1.3 mm rotor, ^{2}H,^{13}C,^{15}N-sample)**
Chemical shift anisotropy (CSA)	Averaged out by global tumbling (< 5 Hz residual)	^{1}H: 4–7 kHz ^{13}CO: 20–23 kHz $^{13}C_{\alpha}/^{13}C_{\beta}$: 4–6 kHz ^{15}N: 11–13 kHz **NB:** fully present in microcrystalline samples, partially averaged out in sedimented samples
J-coupling	^{1}H–^{1}H: 5–15 Hz ^{1}H–^{15}N: 90 Hz	^{1}H–^{1}H: 5–15 Hz ^{1}H–^{15}N: 90 Hz
Dipolar interaction	–	^{1}H–^{1}H: 10–30 kHz ^{1}H–^{15}N: 10–12 kHz ^{13}C–^{15}N: $\sim$ 2 kHz ^{15}N–^{15}N: $<$ 1 kHz ^{13}C–^{13}C: $\sim$ 2 kHz **NB:** dipolar constants scale with the product of the gyromagnetic ratios γ and with r^{-3}.
RDC	^{1}H–^{1}H: 1–10 Hz ^{1}H–^{15}N: 5–20 Hz	–
Mechanism of NOE	Modulation of dipolar couplings by molecular tumbling	Spin diffusion
Linewidths	1–5 Hz	50–150 Hz
^{1}H 90° pulse length	10–15 μs	1 μs

Table A.4: Buffers for protein purification.

Buffer	**Code**	**Composition**
		SH3
AEC A	A	20 mM Tris, pH 8.8 (adjusted with 30% HCl(aq)).
AEC B	B	20 mM Tris, 1 M NaCl, pH 8.5 (adjusted with 30% HCl(aq)).
SEC buffer	S	20 mM citric acid, 150 mM NaCl, pH 3.5 (adjusted with 4 M NaOH(aq)).
Dialysis	D	H_2O, pH 3.5 (adjusted with 30% HCl(aq)).
		TS
T-buffer	T	50 mM Tris–HCl, pH 7.8, 0.1 M NaCl, 1 mM DTT, 5 mM EDTA, 1 mM PMSF, 0.01 mM PLP.
B-buffer	B	50 mM Bicine, pH 9 (using CsOH), 1 mM EDTA, 0.01 mM PLP, 1 mM DTT.
P-buffer	P	0.1 M K_2HPO_4, 0.1 M KH_2PO_4, 5 mM EDTA, 0.01 mM PLP, 170 $g\,L^{-1}$ $(NH_4)_2SO_4$, 1 mM DTT.
Bicine buffer	I	50 mM Bicine, pH 7.8 (adjusted with CsOH).
Crystallization buffer 1	C1	50 mM Cs–bicine buffer, pH 7.8, 14% PEG8000, 3.0 mM spermine.
		PfTrpB
Lysis	L	50 mM KPB, pH 8.0, 20 mM imidazole, 100 mM NaCl, 200 μM PLP, 1 mM PMSF, 1 mM TCEP, 0.02 $mg\,mL^{-1}$ DNase.
Elution	E	50 mM KPB, pH 8.0, 500 mM imidazole, 100 mM NaCl.
NMR buffer	N	50 mM KPB, pH 8.0, 1 mM TCEP.
HEPES buffer	H	20 mM HEPES, pH 7.85.
Crystallization buffer 2	C2	20 mM Na HEPES, pH 7.85, 15–25% PEG3350.
Crystallization buffer 3	C3	0.2 M sodium acetate, 0.1 M Tris, pH 8.5, 30% (w/v) PEG4000.
Crystallization buffer 4	C4	0.2 M $CaCl_2$, 0.1 M sodium acetate, pH 4.3, 15% isopropanol.
		mNTD
Equilibration buffer	Q	50 mM HEPES, pH 8.0, 0.3 M NaCl, 1 mM TCEP, 1 mM PMSF, 10% (w/v) glycerol, 20 mM imidazole.
Elution buffer	U	50 mM HEPES, pH 8.0, 0.3 M NaCl, 1 mM TCEP, 1 mM PMSF, 10% (w/v) glycerol, 0.5 M imidazole.
3C buffer	3C	50 mM HEPES, pH 8.0, 0.3 M NaCl, 1 mM TCEP.
Denaturing buffer	DN	50 mM HEPES, pH 8.0, 0.3 M NaCl, 6 M Gu–HCl, 1 mM TCEP, 1 mM PMSF.
Renaturing buffer	RN	50 mM HEPES, pH 8.0, 0.5 M NaCl, 10% glycerol, 5 mM TCEP, 1 mM PMSF.
NMR buffer	M	20 mM MES, pH 6.0, 0.1 M NaCl, 5 mM TCEP.

Table A.5: Comparison of three cell extract brands used for tackling the H/D exchange problem *(Part 1)* (Chapter 2).

	LB/H_2O	**M9/D_2O**	**10X BioExpress Bacterial Cell Media®/H_2O**	**ISOGRO® (Sigma)/H_2O**	**SILEX® (SILANTES)/H_2O**
Supplements source	–	–	Algal cell hydrolysate	Micro-algal lysate	Chemolithoautotrophic β-proteobacterium *Cupriavidus necator* (syn. *Ralstonia eutropha*)
Features	Fast growth, no isotopic labeling possible	Slow growth due to the presence of D_2O, problem of H/D exchange	Supplied as a liquid concentrate, provides a broad mix of labeled amino acids and cofactors	Supplied as a powder, needs pH adjustment. Provides a broad mix of labeled amino acids and cofactors	Supplied as a powder; ratio and content of amino-acid mixture may differ from ISOGRO and BioExpress due to the bacterial origin of the extract
Labeling	–	U-^{13}C, 98%; U-^{15}N, 98%; U-D, 98% (in exchangeable sites)	U-^{13}C, 98%; U-^{15}N, 98%; U-D, 98%	U-^{13}C, 99%; U-^{15}N, 98%; U-D, 97–99%	U-^{13}C, >98%; U-^{15}N, >98%; U-D, >97%
Supplements per 1 L (as sole source of carbon and nitrogen)	–	–	100 mL	10 g	10 g
Price per 1 L (EUR)	–	–	4150 EUR	1598 EUR	2490 EUR
Amount of glucose per 1 L	–	2 g	–	–	–
Amount of NH_4Cl per 1 L	–	1 g	–	–	–

Table A.5 Comparison of three cell extract brands used for tackling the H/D exchange problem *(Part 2)*

	LB/H_2O	**M9/D_2O**	**10X BioExpress Bacterial Cell Media®/H_2O**	**ISOGRO® (Sigma)/H_2O**	**SILEX® (SILANTES)/H_2O**
Price of glucose per 1 L (EUR)	–	440 EUR	–	–	–
Price of NH_4Cl per 1 L (EUR)	–	112 EUR	–	–	–
D_2O amount	–	Up to 1 L	–	–	–
D_2O price (EUR)	–	925 EUR	–	–	–
Contents	–	ND	30% salts, 3% water, 2% glucose and 65% amino acids/peptides	ND	–
Total price per 1 L $^{2}H/^{13}C/^{15}N$-labelled culture (EUR)	57.4 EUR	2029 EUR	4206 EUR	1654 EUR	2555 EUR

Table A.6: Spectra recorded for PfTrpB samples. Red - DCN labeling, blue – HCN labeling (Chapter 3).

Sedimented PfTrpB	**Crystalline PfTrpB**	**Solution PfTrpB**
HCN – 0.7 mm, 100 kHz DCN – 1.3 mm, 55.5 kHz	DCN – 1.3 mm, 55.5 kHz	Shigemi tube, 45°C
2D hNH 2D hNH 3D hCONH 3D hCONH 3D hCANH	2D hNH 3D hCONH 3D hCANH 4D shCOCANH 4D shCACONH 4D hCACBcaNH	2D HMQC-TROSY 2D HMQC-TROSY 3D HNCO 3D HNcaCO 3D HNcoCA 3D HNCA 3D HNCACB

Table A.7: Assigned chemical shifts (ppm) of PfTrpB (Chapter 3).

Residue number	Residue	H	N	CA	CB	CO
1	MET			57.131	35.551	176.830
2	TRP	9.876	121.087	55.192	31.551	
3	PHE	10.503	124.381	58.336	40.591	174.259
4	GLY	9.041	110.208	50.181		173.129
5	GLU	11.187	127.516	59.025	31.489	173.604
6	PHE	8.978	116.441	60.479	43.481	174.344
7	GLY	9.010	115.047	47.179		174.369
8	GLY	10.225	108.470	48.701		171.556
9	GLN	7.954	116.461	53.586		175.711
10	TYR	9.498	115.550	61.364		169.191
12	PRO			64.639	33.491	176.115
13	GLU	9.337	116.056	60.455	45.040	176.729
14	THR	8.950	112.597	67.463		176.143
15	LEU	7.853	113.392	58.839		176.017
16	VAL	9.266	117.828	63.116	35.145	174.067
17	GLY	9.931	113.541	46.027		171.127
33	GLU			63.011	30.983	177.685
34	GLU	9.238	120.296	61.798		178.376
35	PHE	9.943	121.077	64.902	41.887	175.820
36	ASN	9.015	113.474	60.414	41.970	176.306
37	ARG	9.443	121.013	62.753	31.929	178.903
38	GLN	9.101	118.245	61.980		176.439
39	LEU	9.411	119.722	61.213		177.695
40	ASN	9.321	114.584	59.361	40.702	176.621
41	TYR	8.937	120.905	64.552		178.033
42	TYR	9.445	119.485	61.968		178.668
43	LEU	8.881	121.105	60.612		176.359
52	PRO			62.380		177.136
53	LEU	9.469	121.639	60.170		180.735
54	TYR	9.501	118.736	62.010		177.534
55	TYR	8.250	116.990	60.670	31.924	176.228
59	LEU			60.984		177.461
60	THR	9.190	113.362	70.255	71.723	177.617
61	GLU	9.256	120.839	61.713	31.300	177.142
62	LYS	9.017	119.482	61.709	34.462	177.511
63	ILE	8.646	118.080	66.395	39.307	177.342
64	GLY	8.668	103.598	48.793		173.236
65	GLY	9.632	111.883	46.792		172.367
66	ALA	8.074	126.822	54.987	22.553	175.518
67	LYS	9.084	115.849	61.948		174.960
68	VAL	8.545	118.168	65.183		174.859
69	TYR	9.217	130.130	59.116		173.388
70	LEU	8.845	123.375	56.009		174.038

Continued on next page

Residue number	Residue	H	N	CA	CB	CO
71	LYS	8.726	122.133		35.024	176.678
72	ARG	9.327	119.290	61.835		
76	VAL			64.447		
77	HIS	8.129	117.199	60.577		
78	GLY	11.012	110.115	47.838		172.499
79	GLY	8.389	109.326	46.362		172.460
80	ALA	9.296	123.532	57.338		172.592
81	HIS	9.252				
82	LYS	6.596	117.787	62.243		177.013
83	THR	8.405	114.779			173.077
84	ASN	8.569	116.488	59.825		176.202
85	ASN			63.622		175.367
86	ALA	8.861	128.786	56.108		172.481
87	ILE	7.597	114.838	65.167		170.441
88	GLY	8.664	103.589	49.853		171.971
89	GLN	9.268	114.316	56.418		178.210
95	LEU			60.456		
96	MET			60.024		175.317
97	GLY	8.850	105.866	47.917		174.191
98	LYS	9.139	116.617	57.440		175.423
99	THR	9.540	107.704	63.940		172.886
100	ARG	8.975	124.374	56.848	35.819	172.287
120	LEU			56.671	40.132	174.382
121	LEU	8.467	110.091	56.427		175.634
122	GLY	8.723	105.096	49.401		174.050
123	MET	9.461	117.558	56.861	37.727	174.181
125	VAL	9.038	117.169	62.362		
126	ASP	8.368	124.585	61.029	42.950	179.610
127	ILE	8.557	118.381	61.888	32.080	176.761
128	TYR	8.790	118.283	57.451		178.646
129	MET	9.326	110.854	57.991		176.678
130	GLY	8.700	107.516			
131	ALA	9.593	126.564	58.544		178.716
132	GLU	9.919	115.334	62.098		177.823
133	ASP	8.286	118.852	61.827	32.637	178.599
134	VAL	9.112	121.903	57.784	21.902	179.225
135	GLU			57.147		
143	ARG					174.603
144	MET	8.812	114.813	59.822	34.483	173.122
145	LYS	8.632	116.991	61.138		176.944
146	LEU			57.872	43.337	175.261
147	LEU	8.733	122.982	57.521	43.350	178.215
148	GLY	8.733	105.292	47.231		173.350
149	ALA	8.864	122.204	54.260		175.311
150	ASN	9.269	117.902	54.584		172.870

Continued on next page

Residue number	Residue	H	N	CA	CB	CO
151	VAL	8.890	123.615	64.280	34.322	173.829
152	ILE	10.425	132.314	61.346	40.531	171.903
153	PRO			63.498		176.891
154	VAL	9.826	123.485	65.265	34.309	175.126
155	ASN	9.999	126.955	61.243		179.124
156	SER	8.059	112.900	59.591	67.306	172.685
157	GLY	9.379	106.070	49.755		173.941
158	SER	9.747	121.648	60.660	66.368	172.866
159	ARG	9.465	115.548	59.939		174.204
160	THR	9.798	108.594	62.798		174.461
161	LEU	9.377	123.832	61.284	43.968	177.958
162	LYS	8.630	117.040	63.224		174.722
163	ASP	10.232	125.231	57.344		174.461
164	ALA	8.089	119.804	52.305	25.817	173.162
165	ILE	9.114	119.720	56.936		172.608
166	ASN	9.340	120.590	60.039		174.888
167	GLU	8.551	118.115			
172	TRP			62.050		176.784
173	VAL	8.736	111.481	68.857		175.057
174	ALA	8.288	117.029	55.853		179.827
175	THR	9.102	107.419			176.632
176	PHE	8.743	120.821	62.536		174.713
177	GLU	9.016	122.195	62.607	31.452	175.431
184	GLY			46.991		172.833
185	SER	7.736	110.738	58.479	65.790	174.182
186	VAL	8.545	113.472	61.832		174.183
187	VAL	8.685	111.321	60.975		171.141
188	GLY	8.293	101.557	45.449		170.827
193	PRO					176.979
194	THR			68.952		172.839
195	ILE	9.591	124.832	64.922		176.383
196	VAL	8.876	115.118	62.301		
221	VAL			66.356		176.158
222	ILE	10.146	121.632	62.791		178.174
223	VAL	7.608	113.522	61.892	33.655	175.795
224	ALA	7.648	119.778	58.071		178.271
225	CYS	8.831	117.677	63.408	35.518	178.577
227	GLY			43.806		176.205
228	GLY	9.323	111.973	46.909		172.032
229	GLY	8.732	102.034	49.209		172.026
230	SER	7.983	112.087	56.117	71.386	171.436
231	ASN	8.811	114.814	55.362	42.104	173.171
232	ALA			56.149		
233	MET	8.899	112.053	58.265	31.536	178.231
234	GLY	9.118	103.622	50.753		173.107

Continued on next page

Residue number	Residue	H	N	CA	CB	CO
235	ILE	8.296	115.162	65.370	41.737	173.820
236	PHE	9.824	116.382	59.554		179.203
237	TYR	9.608	115.092	67.678		174.117
247	LEU			56.993		175.759
248	VAL	8.212	120.168	70.276	33.495	
249	GLY	8.198	105.229	50.802		174.045
250	VAL		122.016	63.032		176.143
252	ALA			55.309		176.254
253	GLY	10.457	112.668	47.268		173.131
254	GLY	9.881	107.436	49.106		175.810
255	LYS	9.788	125.455	58.394		175.226
256	GLY	9.841	109.360	46.787		176.151
257	LEU	9.989	127.047	55.898		172.842
268	ALA			54.412		176.251
269	GLY	8.371	107.978	47.456		171.552
270	GLN	9.557	115.442	56.627		173.189
271	VAL	9.296	122.426	65.697		176.308
272	GLY	10.118	113.695	47.632		169.187
273	VAL	8.834	117.861	62.880		175.976
274	SER	9.267	117.761			
275	HIS			58.059		173.890
276	GLY	9.388	99.000	48.618		172.619
277	MET			56.218		170.437
278	LEU	10.582	123.307	57.253		177.261
279	SER	9.911	118.368	58.281		172.376
280	TYR	8.734	116.558	60.996	42.438	
284	ASP				40.494	
285	GLU			61.491	32.280	176.275
286	GLU	9.082	117.359		32.188	
287	GLY	8.936	107.702	48.186		173.395
288	GLN	9.244	118.968	56.788	31.154	172.713
289	ILE	9.377	119.003	59.500	42.590	174.912
290	LYS	9.490	116.832	61.403		172.738
294	SER				67.162	
318	GLU			64.680		
319	TYR	10.378	117.913	60.251		177.132
320	VAL	9.014	114.815	62.272	31.773	178.355
321	ALA	9.258	121.563	57.475	20.307	178.279
322	VAL	9.820	117.838	64.960	40.173	175.781
323	THR	8.892	115.132	62.116		176.531
377	ASP			56.381	40.725	173.277
378	LEU	8.557	121.261	56.884	43.820	174.377
379	ASP	8.741	122.153	59.082	35.145	174.649
380	ILE	9.449	123.712	64.340	36.359	173.192
381	VAL	8.057	120.048	67.323		174.752

Continued on next page

Residue number	Residue	H	N	CA	CB	CO
382	LEU	9.144	113.179	59.422		177.243
383	LYS	8.926	123.100	58.545		177.295
384	ALA			58.544		
385	SER				63.790	176.375

Table A.8: N-terminal domain annotations of vertebrate cGAS proteins retrieved using CD-Search.

Protein domain annotation	**Domain family accession number**	**Lowest E-value in the orthologous group**	**Species**
DNA polymerase subunits gamma and tau	PRK07764, PRK07003, PRK12323	6.43E-11	*Pan troglodytes, Pan paniscus, Homo sapiens, Gorilla gorilla, Pongo abelii, Macaca mulatta, Physeter macrocephalus, Capra hircus, Bos taurus, Camelus dromedarius, Sus scrofa, Myotis yumanensis, Canis lupus familiaris, Felis catus, Loxodonta africana, Ornithorhynchus anatinus, Crocodylus porosus*
Fragment of UL36 large tegument protein	PHA03247	1.44E-08	*Rattus norvegicus, Mus musculus, Cricetulus griseus, Tursiops truncatus, Orcinus orca*
VP22 tegument protein	PHA03381	9.59E-10	*Marmota marmota marmota, Salminus brasiliensis*
ICP4 transcriptional regulator protein	PHA03307	1.1E-14	*Cricetulus griseus, Balaenoptera musculus, Eptesicus fuscus, Hipposideros armiger, Pteropus vampyrus, Dermochelys coriacea, Gekko japonicus, Microcaecilia unicolor, Rhinatrema bivittatum*
Epstein–Barr early nuclear antigen 3B (EBNA-3B)	PHA03378	6.23E-11	*Desmodus rotundus, Mustela putorius furo, Monodelphis domestica*
g150 protein from Epstein–Barr virus (BDLF3)	PHA03255	5.61E-05	*Phyllostomus discolor*
FhaB hemagglutinin protein from toxin–antitoxin system of *Bordetella pertussis*	COG3210	4.53E-17	*Tachyglossus aculeatus*
Merozoite erythrocyte binding protein 2 from *Plasmodium* (MAEBL)	PTZ00121	4.91E-14	*Anolis carolinensis, Rhineura floridana, Xenopus laevis*
60S ribosomal L19-like protein	PTZ00436	3.48E-06	*Eretmochelys imbricata, Pelodiscus sinensis*
Transcription termination factor Rho	PRK12678	5.94E-07	*Nannorana parkeri, Microcaecilia unicolor*
Pneumococcal surface protein PspC	cl41463	3.11E-06	*Pemptheris kluzingeri, Sebastes umbrosus*
Provisional adenylate kinase	PRK13808	4.23E-06	*Branchiostoma floridae*

Table A.9: DNA oligonucleotides used in cGAS NTD study (Chapter 4).

Sample	**DNA length**	**DNA sequence**	**GC content, %**
Apo	–	–	–
Random15	15	CGA CGC TAG CGT CGA	66.67
Satellite	20	AAT GGA ATG GAA TGG AAT GG	40
Telomeric	24	TTA GGG TTA GGG TTA GGG TTA GGG	50
AT-rich	24	ATT TTT ACA TTT TTA CAT TTT TAC	12.5
G3-YSD	26	GGG TAT ATA TAT ATA TAT ATA TAG GG	23.08
Random42	42	CGA CGC TAG CGT CGC GAC CGT AGC GTC GCG ACG CTA GCG TCG	71.43
ISM	50	GGA GGC GGA GCT TGC AGT GAG CGA AAA TCG CAC CAC TGC ACT CCA GCC TG	62
HSV	57	GAC ACG ATG CGA TAA AAT CTG TTT GTA AAA TTT ATT AAG GGT ACA AAT TGC CCT AGC	35.09
P-ODN	23	TCG TCG ACG ATC GGC GCG CGC CG	78.26

Appendix B

Supplementary Figures

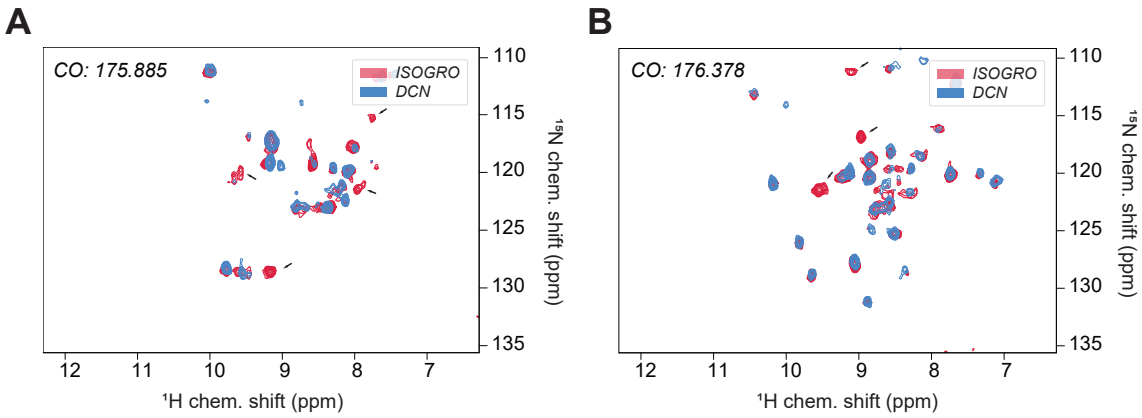


Figure B.1: Representative projections from hCONH spectrum recorded for TS in perdeuterated (DCN) and back-exchanged (ISOGRO®) samples.

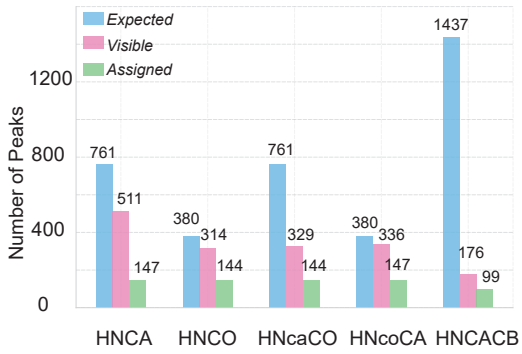


Figure B.2: Statistics on expected, visible, and assigned peaks in solution NMR experiments recorded for PfTrpB^{2B9}.

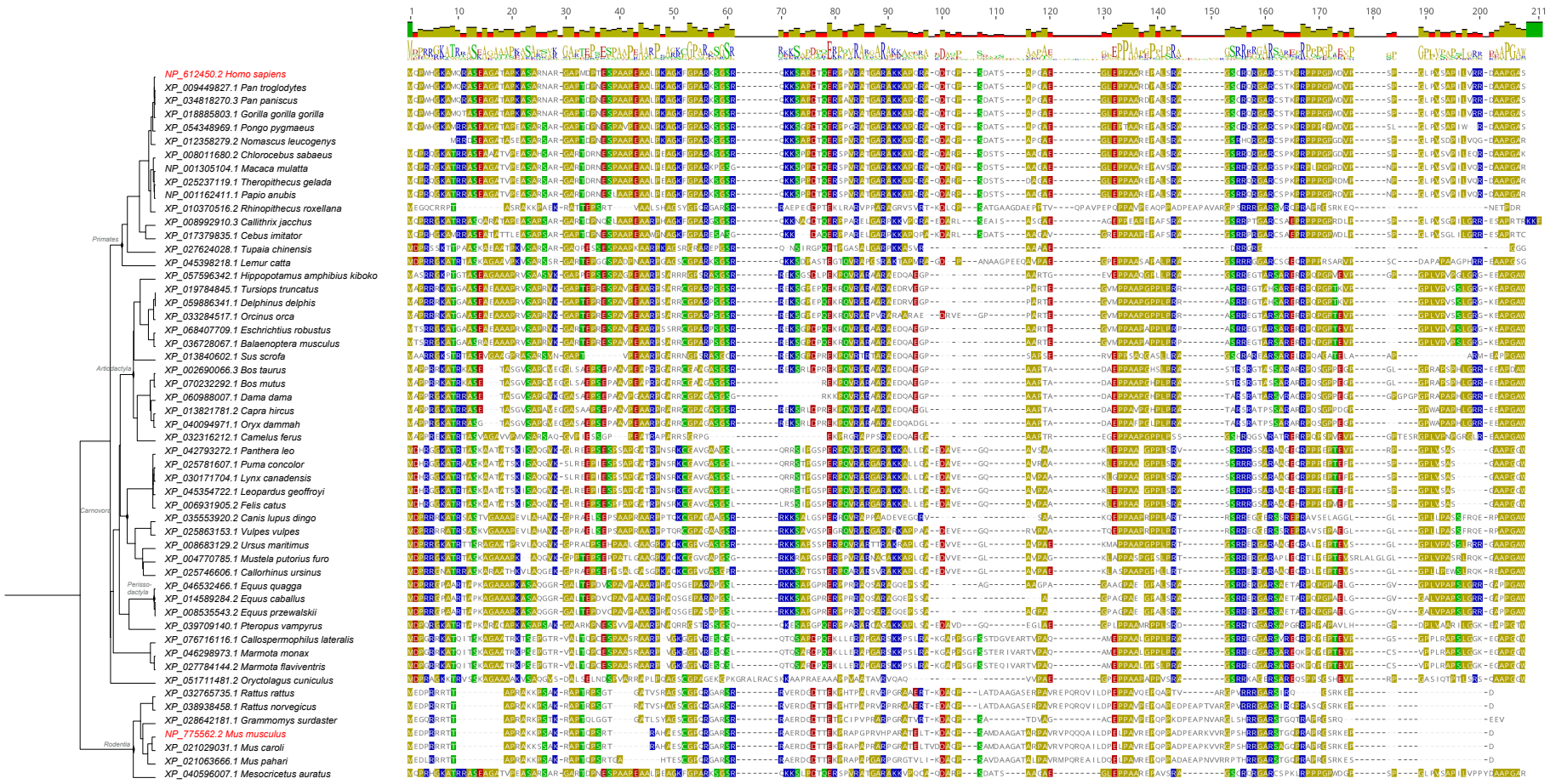


Figure B.3: NTD conservation across vertebrates

Project Number	Project Templatename	Project Targetvector	Project Targetname	Amplification Expected Size(bp)	Project PCR Status	nPCR Comm	Screen-PCR ning Com Expected Size(bp)		Sequencing Sequencing Comment (public)	Expressionte st Module 2	Expressionte st Expected Size(Da)	Expressiontest 37°C	Expressiontest 20°C	Expressiontest AI25°C	Expressiontest Expressiontest Comment (public)
155_46	hcGAScore	pOPIN(c)Cherry	pOPIN(c)Cherry_homo_cGACcore_4	1140	X	ok	1425	ok	ok	Yes	73062	0	0	0	
155_58	hcGAScore	pOPIN(n)Cherry	pOPIN(n)Cherry_homo_cGACcore_6	1140	X	ok	1506	ok	ok, silent mutation D289D	Yes	72845	0	0	0	
155_64	hcGAScore	pOPIN(n)EGFP	pOPIN(n)EGFP_homo_cGACcore_7	1140	X	ok	1473	ok	ok	Yes	73064	0	0	1	additional band at 33kDa
	hcGAScore	pOPIN(c)EGFP	pOPIN(c)EGFP_homo_cGACcore_5	1140	X	ok	1511	ok	deletion in rev primer, frame shift	Yes	73362	N/A	N/A	N/A	N/A
	hcGAScore	pOPIN-His-GST	pOPIN-His-GST_homo_cGACcore_1	1140	X	ok	1381	ok	ok	Yes	72011	0	0	0	
	hcGAScore	pOPIN-His-MBP	pOPIN-His-MBP_homo_cGACcore_2	1140	X	ok	1461	ok	no fwd sequencing result	Yes	86825	0	0	0	
	hcGAScore	pOPIN-His-SUMO	pOPIN-His-SUMO_homo_cGACcore_3	1140	X	ok	1482	ok	ok	Yes	56486	0	1	0	MW deviation! 63kDa
155_47	hcGASfl.opti	pOPIN(c)Cherry	pOPIN(c)Cherry_homo_cGASfl.opti_4	1563	X	ok	1848	ok	ok	Yes	87613	0	0	0	
155_59	hcGASfl.opti	pOPIN(n)Cherry	pOPIN(n)Cherry_homo_cGASfl.opti_6	1563	X	ok	1929	ok	mutation R328C	Yes	87396	0	0	0	total lysate 37°C: unspez. band (air bubbel)
155_65	hcGASfl.opti	pOPIN(n)EGFP	pOPIN(n)EGFP_homo_cGASfl.opti_7	1563	X	ok	1896	no visible	ok	Yes	87615	0	0	0	
	hcGASfl.opti	pOPIN(c)EGFP	pOPIN(c)EGFP_homo_cGASfl.opti_5	1563	X	ok	1934	ok	mutation P100L	Yes	87913	0	0	2	MW deviation! 67kDa
	hcGASfl.opti	pOPIN-His-GST	pOPIN-His-GST_homo_cGASfl.opti_1	1563	X	ok	1804	no visible	silent mutation L404L	Yes	86562	0	0	0	total lysate 20°C: unspez. band (air bubbel)
	hcGASfl.opti	pOPIN-His-MBP	pOPIN-His-MBP_homo_cGASfl.opti_2	1563	X	ok	1884	no visible	ambiguous sequencing result (fwd) - overlap, rev ok +MUTATIONS	Yes	101376	0	0	0	
	hcGASfl.opti	pOPIN-His-SUMO	pOPIN-His-SUMO_homo_cGASfl.opti_3	1563	X	ok	1905	no visible	ok	Yes	71037	0	0	0	
155_48	hcGASntd	pOPIN(c)Cherry	pOPIN(c)Cherry_homo_cGASntd_4	423	X	ok	708	ok	ok, complete sequence coverage (rev result), fwd bad seq result	Yes	43500	0	0	0	
155_60	hcGASntd	pOPIN(n)Cherry	pOPIN(n)Cherry_homo_cGASntd_6	423	X	ok	789	ok	ok, complete sequence coverage	Yes	43282	0	0	0	
155_66	hcGASntd	pOPIN(n)EGFP	pOPIN(n)EGFP_homo_cGASntd_7	423	X	ok	756	ok	ok, complete sequence coverage	Yes	43502	0	0	0	
	hcGASntd	pOPIN(c)EGFP	pOPIN(c)EGFP_homo_cGASntd_5	423	X	ok	794	ok	ok, complete sequence coverage	Yes	43799	0	0	0	
	hcGASntd	pOPIN-His-GST	pOPIN-His-GST_homo_cGASntd_1	423	X	ok	664	no visible	gap - AA348 - AA376 missing	Yes	42448	1	1	0	MW deviation! 49kDa, additional band at 28+37kDa
	hcGASntd	pOPIN-His-MBP	pOPIN-His-MBP_homo_cGASntd_2	423	?	weak PCR prc	744	no visible	ok, complete sequence coverage	Yes	57262	0	1	1	MW deviation! 67kDa, additional band at 44kDa (MBP)
	hcGASntd	pOPIN-His-SUMO	pOPIN-His-SUMO_homo_cGASntd_3	423	X	ok	765	ok	mutation: R162H	Yes	26924	0	0	0	eluate Ai: (49kDa): unspez. band (air bubbel)
155_49	mcGAScore	pOPIN(c)Cherry	pOPIN(c)Cherry_mus_cGAScore_4	1083	X	ok, unspec. b	1368	ok	deletion in rev primer, frame shift	Yes	71285	N/A	N/A	N/A	N/A
155_61	mcGAScore	pOPIN(n)Cherry	pOPIN(n)Cherry_mus_cGAScore_6	1083	X	ok	1449	ok	insertion in fwd primer, frame shift	Yes	71068	N/A	N/A	N/A	N/A
155_67	mcGAScore	pOPIN(n)EGFP	pOPIN(n)EGFP_mus_cGAScore_7	1083	X	ok	1416	ok	mutation L618F	Yes	71287	0	0	0	
	mcGAScore	pOPIN(c)EGFP	pOPIN(c)EGFP_mus_cGAScore_5	1083	X	ok	1454	ok	ok, silent mutation: see 155_18	Yes	71585	0	0	0	
	mcGAScore	pOPIN-His-GST	pOPIN-His-GST_mus_cGAScore_1	1083	X	ok	1324	ok	ok, silent mutation, see 155_18	Yes	70234	0	0	0	total lysate 20°C: unspez. band (air bubbel)
	mcGAScore	pOPIN-His-MBP	pOPIN-His-MBP_mus_cGAScore_2	1083	X	ok	1404	ok	silent mutations: L421L, F477F and ambiguous sequencing result - overlap	Yes	85048	0	0	0	
	mcGAScore	pOPIN-His-SUMO	pOPIN-His-SUMO_mus_cGAScore_3	1083	X	ok	1425	ok	mutation: R299M and silent mutation (see 155_18)	Yes	54709	0	0	1	MW deviation! 62kDa
155_50	mcGASfl	pOPIN(c)Cherry	pOPIN(c)Cherry_mus_cGASfl_4	1518	X	ok	1803	ok	ok, silent mutation: see 155_35	Yes	86993	3	3	3	MW deviation! 65kDa
155_62	mcGASfl	pOPIN(n)Cherry	pOPIN(n)Cherry_mus_cGASfl_6	1518	X	ok	1884	ok	deletion in fwd primer, frame shift	Yes	86775	N/A	N/A	N/A	N/A
155_68	mcGASfl	pOPIN(n)EGFP	pOPIN(n)EGFP_mus_cGASfl_7	1518	X	ok	1851	ok	deletion in fwd primer, frame shift	Yes	86995	N/A	N/A	N/A	N/A
	mcGASfl	pOPIN(c)EGFP	pOPIN(c)EGFP_mus_cGASfl_5	1518	X	ok	1889	ok	ok, silent mutation: see 155_35	Yes	87292	0	0	0	
	mcGASfl	pOPIN-His-GST	pOPIN-His-GST_mus_cGASfl_1	1518	X	ok	1759	ok	ok	Yes	85941	0	0	0	
	mcGASfl	pOPIN-His-MBP	pOPIN-His-MBP_mus_cGASfl_2	1518	X	ok	1839	ok	deletion in fwd primer, frame shift	Yes	100755	N/A	N/A	N/A	N/A
	mcGASfl	pOPIN-His-SUMO	pOPIN-His-SUMO_mus_cGASfl_3	1518	X	ok	1860	additional	deletion in fwd primer, frame shift	Yes	70417	N/A	N/A	N/A	N/A
155_51	mcGASntd	pOPIN(c)Cherry	pOPIN(c)Cherry_mus_cGASntd_4	435	X	ok	720	ok	deletion in rev primer, frame shift	Yes	44656	N/A	N/A	N/A	N/A
155_63	mcGASntd	pOPIN(n)Cherry	pOPIN(n)Cherry_mus_cGASntd_6	435	X	ok	801	ok	ok, silent mutation see 155_20 and G388G	Yes	44439	0	0	0	
155_69	mcGASntd	pOPIN(n)EGFP	pOPIN(n)EGFP_mus_cGASntd_7	435	X	ok	768	ok	ok, silent mutation (see155_20), complete sequence coverage	Yes	44658	0	3	1	MW deviation!59kDa, additional band at 31kDa
	mcGASntd	pOPIN(c)EGFP	pOPIN(c)EGFP_mus_cGASntd_5	435	X	ok	806	ok	ok, silent mutation: see 155_20, complete sequence coverage	Yes	44955	0	0	3	MW deviation!60kDa
	mcGASntd	pOPIN-His-GST	pOPIN-His-GST_mus_cGASntd_1	435	X	ok	676	ok	ok, silent mutation, see 155_20	Yes	43605	0	0	0	
	mcGASntd	pOPIN-His-MBP	pOPIN-His-MBP_mus_cGASntd_2	435	X	ok	756	ok	ok, silent mutation R517R	Yes	58419	3	3	3	MW deviation! 68kDa, additional band at 54kDa
	mcGASntd	pOPIN-His-SUMO	pOPIN-His-SUMO_mus_cGASntd_3	435	X	ok	777	ok	mutation: A153V and silent mutation, see 155_20	Yes	28080	0	0	1	

Figure B.4: Results of human and murine cGAS NTD, cGAS core, and cGAS full-length gene cloning into pOPINE vectors encoding various fusion partners. Columns 'Expressiontest' at various temperatures show relative amount of protein produced from transforming *E.coli* RIPL (DE3) pLysS cells with designated pOPINE vectors, with N/A standing for no information due to the errors in cloning, 0 standing for $\prec$5 ng/μl expression yield, 1 – 5-50 ng/μl, 2 – 50-250 ng/μl, and 3 - $\succ$250 ng/μl. All vectors with serious errors during cloning such as frameshift mutations, indels, or lacking sequencing results were discarded; vectors carrying silent mutations were considered suitable for the study.

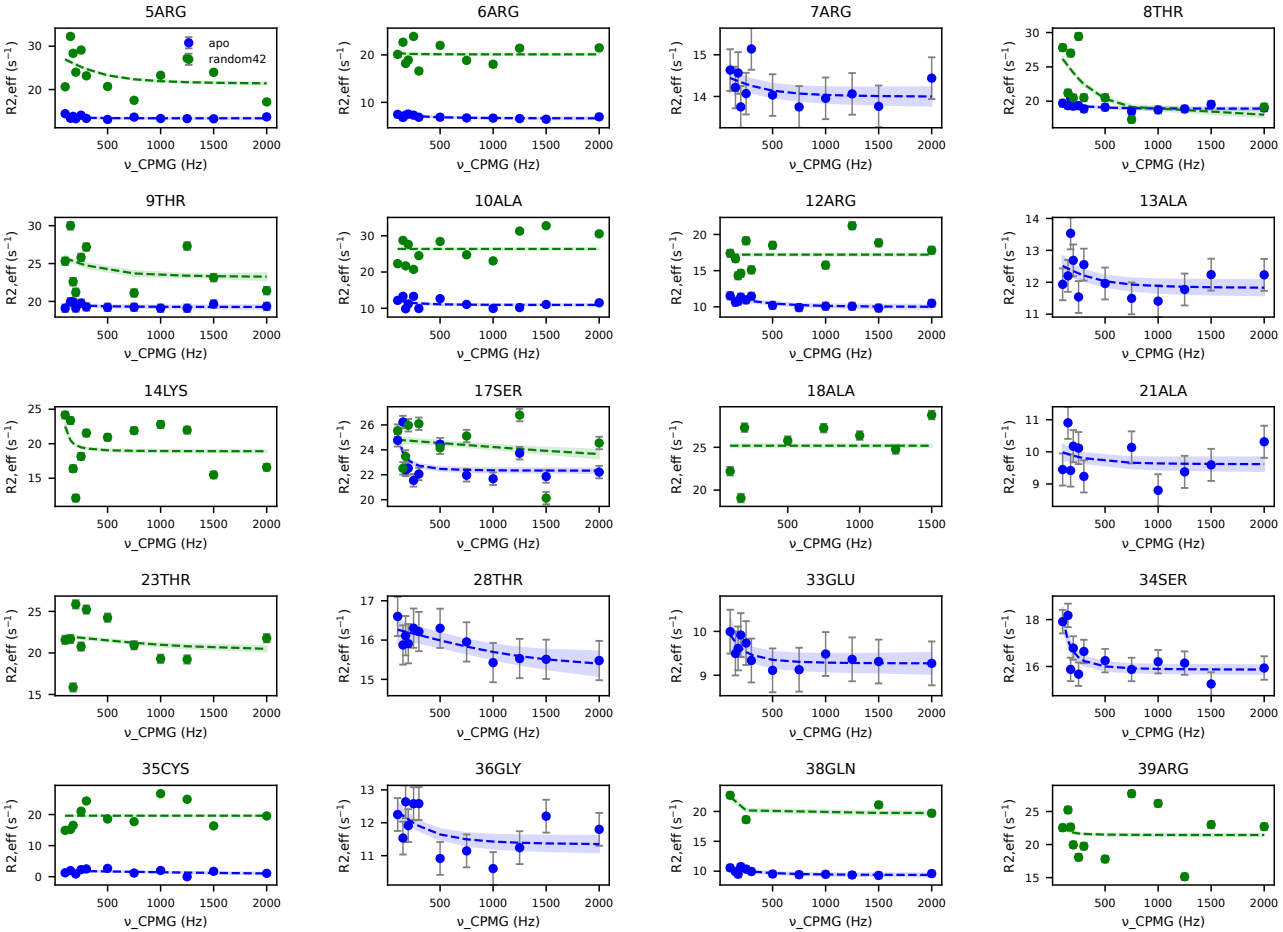
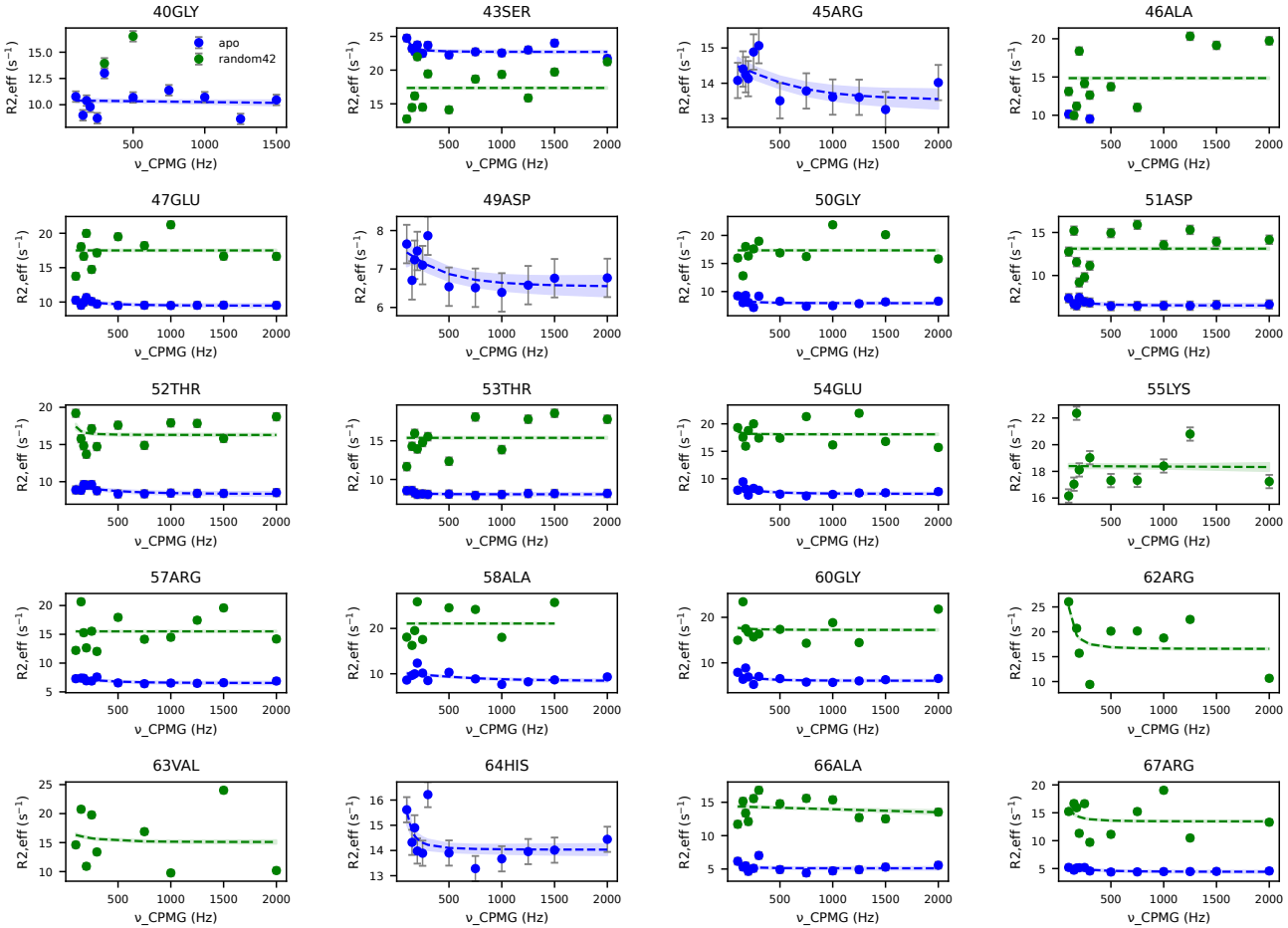
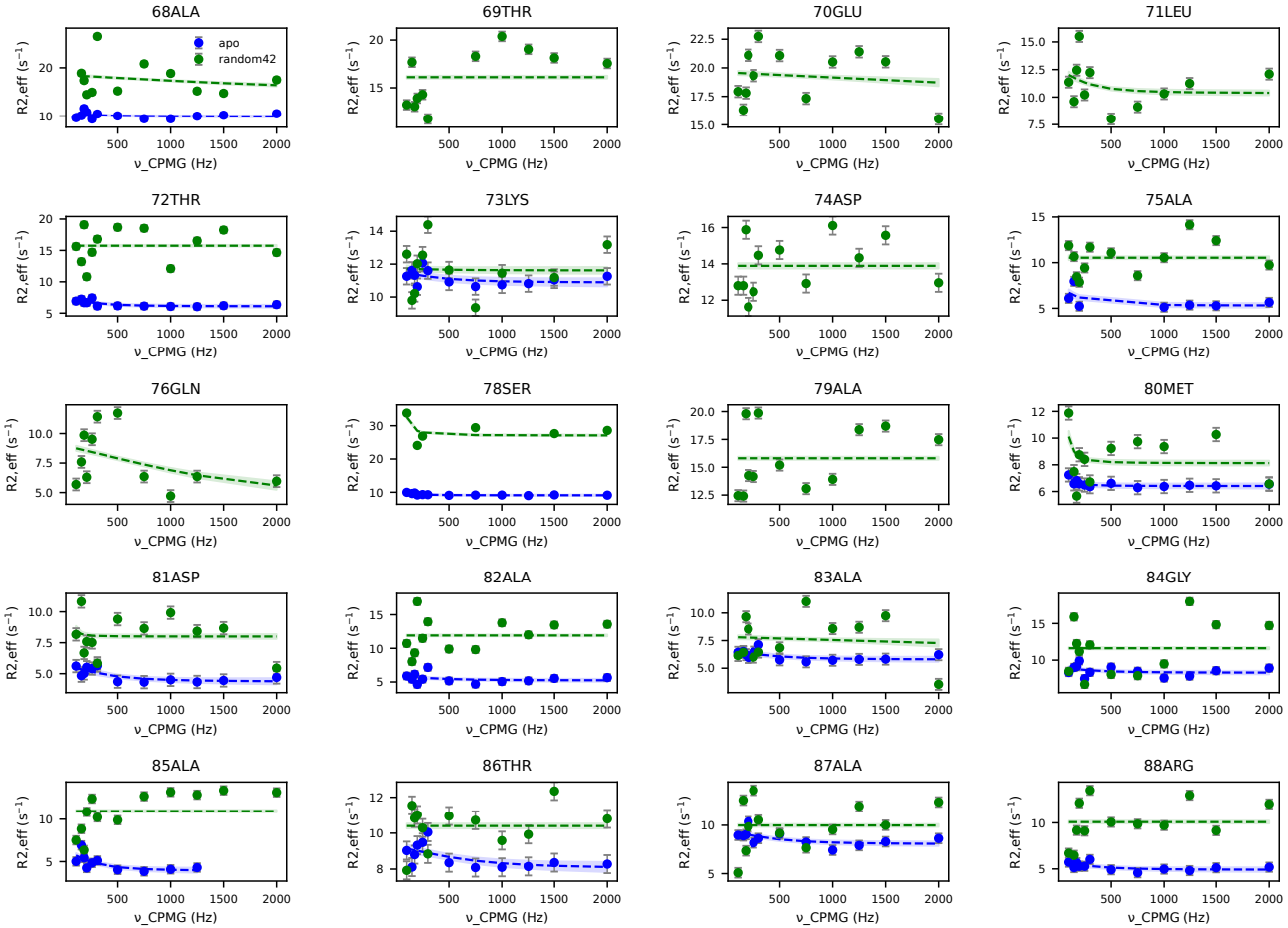
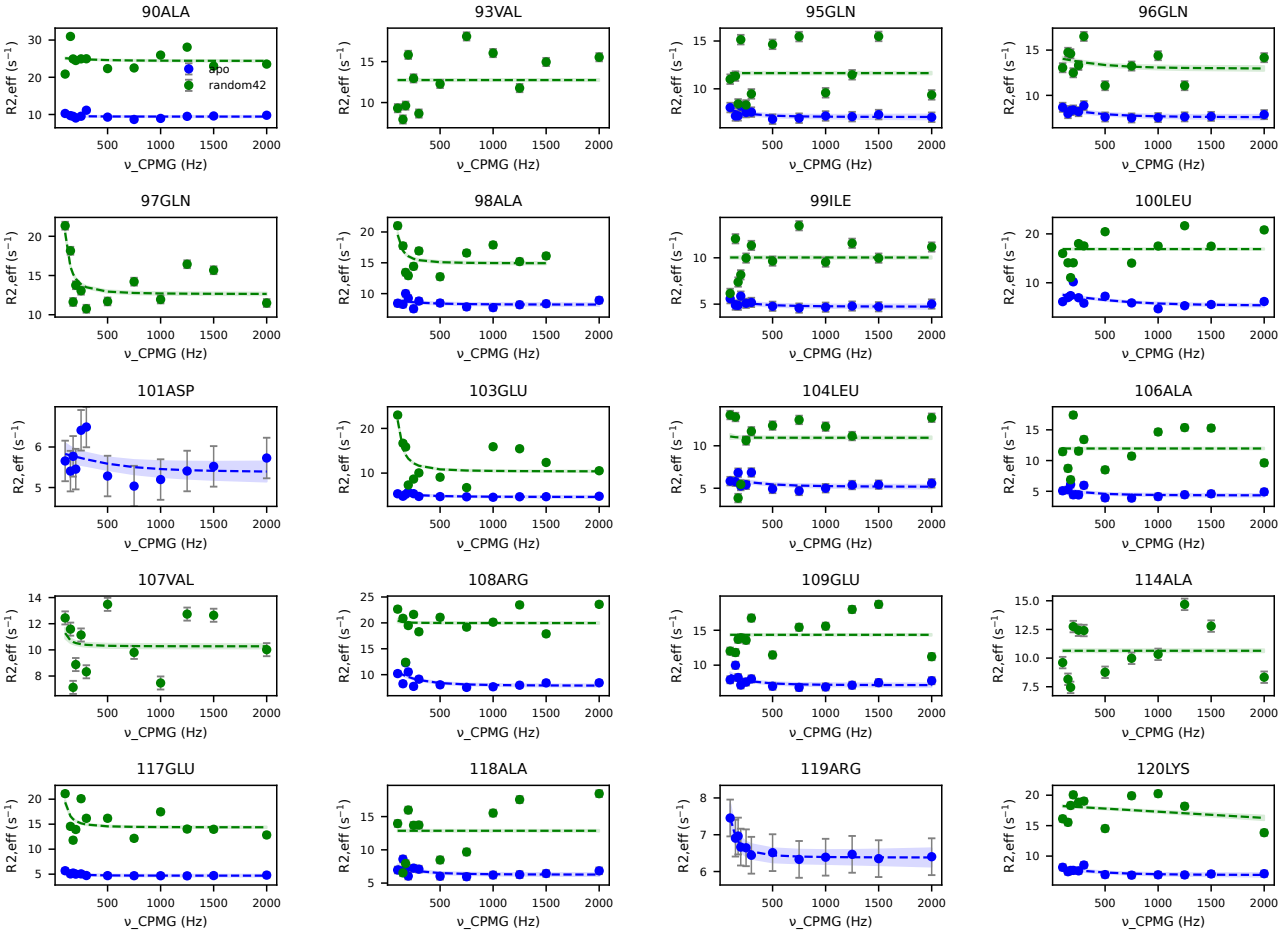
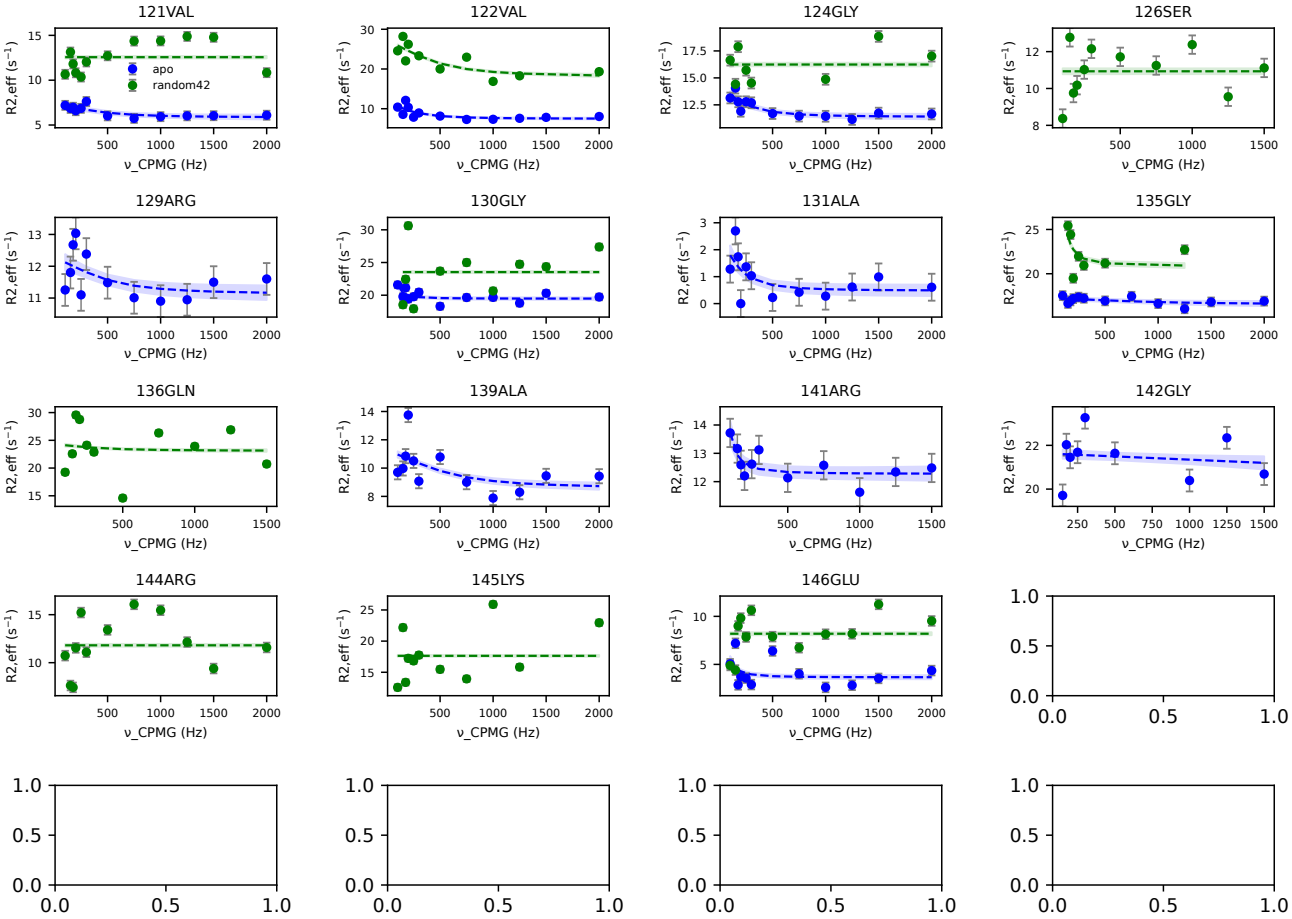


Figure B.5: CPMG profiles of cGAS NTD in apo form (blue) and with an excess of random42 DNA (green).


40GLY
apo
random42
43SER
45ARG
46ALA
47GLU
49ASP
50GLY
51ASP
52THR
53THR
54GLU
55LYS
57ARG
58ALA
60GLY
62ARG
63VAL
64HIS
66ALA
67ARG
R2,eff (s⁻¹)
ν_CPMG (Hz)


68ALA
apo
random42
69THR
70GLU
71LEU
72THR
73LYS
74ASP
75ALA
76GLN
78SER
79ALA
80MET
81ASP
82ALA
83ALA
84GLY
85ALA
86THR
87ALA
88ARG
R2,eff (s⁻¹)
ν_CPMG (Hz)


90ALA
93VAL
95GLN
96GLN
97GLN
98ALA
99ILE
100LEU
101ASP
103GLU
104LEU
106ALA
107VAL
108ARG
109GLU
114ALA
117GLU
118ALA
119ARG
120LYS
R2,eff (s⁻¹)
ν_CPMG (Hz)
apo
random42


121VAL
122VAL
124GLY
126SER
129ARG
130GLY
131ALA
135GLY
136GLN
139ALA
141ARG
142GLY
144ARG
145LYS
146GLU
apo
random42
R2,eff (s⁻¹)
ν_CPMG (Hz)