University of Louisville

ThinkIR: The University of Louisville's Institutional Repository

Electronic Theses and Dissertations

8-2022

Applications of nuclear magnetic resonance spectroscopy: from drug discovery to protein structure and dynamics.

Mark Vincent C. dela Cerna University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

Part of the Biochemistry Commons, Molecular Biology Commons, Other Biochemistry, Biophysics, and Structural Biology Commons, and the Structural Biology Commons

Recommended Citation

dela Cerna, Mark Vincent C., "Applications of nuclear magnetic resonance spectroscopy: from drug discovery to protein structure and dynamics." (2022). *Electronic Theses and Dissertations.* Paper 3946. Retrieved from https://ir.library.louisville.edu/etd/3946

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact thinkir@louisville.edu.

APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY: FROM DRUG DISCOVERY TO PROTEIN STRUCTURE AND DYNAMICS

by

Mark Vincent Carreon dela Cerna Bachelor of Science, Ateneo de Manila University 2012 Master of Chemistry, Ateneo de Manila University 2014 Master of Science, University of Louisville 2019

A Dissertation Submitted to the Faculty of the School of Medicine of the University of Louisville in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy In Biochemistry and Molecular Genetics

Department of Biochemistry and Molecular Genetics University of Louisville Louisville, Kentucky

August 2022

APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY: FROM DRUG DISCOVERY TO PROTEIN STRUCTURE AND DYNAMICS

by

Mark Vincent Carreon dela Cerna Bachelor of Science, Ateneo de Manila University 2012 Master of Chemistry, Ateneo de Manila University 2014 Master of Science, University of Louisville 2019

A Dissertation Approved on

July 18, 2022

By the following Dissertation Committee

Dissertation Director T. Michael Sabo, Ph.D.

Dissertation Co-Director Brian Clem, Ph.D.

Carolyn Klinge, Ph.D.

Jessica Blackburn, Ph.D.

Corey Watson, Ph.D.

Ying Li, Ph.D.

DEDICATION

To goals not yet attained. To questions left unanswered. To stories untold and voices unheard.

To Alicia. To my mom, Chona. To my brother, Richard. To Rita and Junior.

Isigaw mo sa hangin, tumindig at magsilbing liwanag, liwanag sa dilim Harapin mong magiting ang bagong awitin Ikaw ang LIWANAG SA DILIM (Liwanag sa Dilim, Rivermaya, 2005)

ACKNOWLEDGEMENTS

"We're all stories in the end. Just make it a good one, eh? Cause it was, you know. It was the best." --Eleventh, Doctor Who

Ah, who am I kidding? This part is <u>far</u> from the best. Nonetheless, it is a chapter closed, and no doubt, made bearable by mentors, colleagues, friends, and family.

First, to my mentors: Donghan for allowing me to pursue my passions and interests, for showing me the 'behind the scenes' of an academic lab, and for giving me freedom to teach and mentor my own students while working on my PhD; Mike for teaching me 75% (my estimate!) of the NMR I know and for stepping in to be my mentor; Jessi for adopting me as a lab member (*kinda*), for showing me that excellent mentorship knows no distance, for showing her excitement for my victories, big or small, and for being the best collaborator in my short career so far.

To more mentors: Dungeng Peng for always wanting the best for me and for believing in me from the very beginning; Elias Fernandez for offering support during my stressful research transition; David Ban for pushing me into the deep end on my first project (analyzed RDCs before I ran my first HSQC!) and for my first beer in Louisville; Joe Burlison for making organic chemistry *sound cool* and for listening to my random rants; Brian Clem for helping us (Donghan, Mike, and I) navigate the ins-and-outs of the graduate program; Chi Li for exposing me to metabolism and allowing me to apply NMR to this field; Hassane Mchaourab for giving me the confidence to dream big; Manuel Ascano for teaching me that it is okay to fail and that it is even <u>better to fail fast</u>; Walter Chazin for sharing the story of his own PhD journey which has since served as a motivation to succeed in mine.

To the rest of my dissertation committee: Carolyn Klinge, Ying Li, and Corey Watson, for the questions and the guidance, and for making it so easy to schedule committee meetings and seminars (*this applies to all seven of you!*).

Next, to my colleagues. Bellarmine's Department of Chemistry: specifically, Pat Holt for welcoming me into his department and for providing me with increasing teaching responsibilities over the years; Theresa Hahn for sharing her wisdom and allowing me to discover myself as an educator; Amanda Krzysiak for trusting me with mentoring BMB students and for writing a letter that might have just changed my life; Colleagues at the Cancer Center and the Biochemistry Department: Nazimmudin Khan, Will Holmes, Sam Arumugam, Lynn DeeLeuw, Jessica Cannon, Velma Shang, Cassy Luckett, Rob Monsen, Shin-Je Ghim. Former colleagues at Vanderbilt: Smriti Mishra, Sanjay Mishra, Kevin Jagessar, Derek Claxton, Richard Stein. Everyone at the Kentucky Academy of Sciences. **My Bellarmine group** for being the most awesome trainees one could ever have: Jacob Santana, Tristan Whalen, Chase Yost, Kate Weafer, Olivia Paris, Yvonne Ndoricimpa, Rahim Shalash, Olivia Gamsky, Kat Tikhonko. <u>Go Knights!!</u>

To friends, near and far, for providing a support system, for sharing in my successes and failures, and for, generally, sticking around when all I can talk about

V

are structural biology and proteins and how all I want in life is to run my own research lab. I want to especially single out: Allie Fuller, Francisco Rodriguez, Emily Duderstadt, Lara Marayag, Paul Sanchez, Raian Razal.

Finally, to my family. My mom, for *everything*, and especially for enduring almost **<u>25 years</u>** of me in school. My brother Richard, aunts Cynthia and Daisy, uncles Rollie, Dan, and Bong, and cousins RA, Chel, Danna, and Ryan. Rita and Junior Brinegar, for essentially providing me a real *home* here in Louisville.

To our kitties, Mikah, Wednesday, Nox, Lydia, and Clove, and their kitty cousins Minerva, Bitty, Eisa, Beyla, Sassy and doggy cousin Precious, because they contribute to everyone's overall happiness.

Alicia for the love and support, for giving me a reason not to *quit* when I think life is not worth living, and for making life, really, worth living. Also, thank you for sharing my dream: *Now, we got a lab and can actually do whatever we want! Time to* ACTUALLY *work.*

I am also grateful to the University of Louisville and the Integrated Programs in Biomedical Sciences for the financial support, especially to the Iva W. Homberger Scholarship Fund and the BMG department for supporting the final years of my PhD studies.

ABSTRACT

APPLICATIONS OF NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY: FROM DRUG DISCOVERY TO PROTEIN STRUCTURE AND DYNAMICS

Mark Vincent Carreon dela Cerna

July 18, 2022

The versatility of nuclear magnetic resonance (NMR) spectroscopy is apparent when presented with diverse applications to which it can contribute. Here, NMR is used i) as a screening/ validation tool for a drug discovery program targeting the Phosphatase of Regenerating Liver 3 (PRL3), ii) to characterize the conformational heterogeneity of p53 regulator, Murine Double Minute X (MDMX), and iii) to characterize the solution dynamics of guanosine monophosphate kinase (GMPK).

Mounting evidence suggesting roles for PRL3 in oncogenesis and metastasis has catapulted it into prominence as a cancer drug target. Yet, despite significant efforts, there are no PRL3 small molecule inhibitors currently in clinical trials. This work combines screening of an FDA-approved drug panel and the identification of binders by protein-observed NMR. FDA-approved drugs salirasib and candesartan were identified as potent inhibitors in *in vitro* inhibition and migration assays while a weak inhibitor, olsalazine, was identified by NMR as the first small molecule inhibitor to directly bind PRL3. NMR was also used to validate the binding of additional compounds identified as experimental PRL3 inhibitors. Thienopyridone, a potent experimental inhibitor, did not show direct binding to PRL3 but instead inhibited phosphatase activity via redox mechanism. NMR also revealed that other experimental inhibitors did not engage PRL3. Thus, there remains a need to identify potent PRL3-directed inhibitors. Meanwhile, molecular modeling revealed a putative druggable site that has not been thoroughly explored before. The current study provides some scaffolds such as candesartan and particularly, olsalazine, the only binder identified, that could be the starting point of further drug discovery efforts, as well as a putative site that can be targeted *in silico*.

MDMX, a negative regulator of p53, is another important therapeutic target in cancer, along with the homologous protein, MDM2. Inhibitors that block the MDM2p53 interaction have been identified and despite similarities in the binding site of these homologous proteins, these inhibitors are ineffective against MDMX. It is hypothesized that the flexibility of MDMX contributes to this significant difference in response to inhibitors, despite comparable affinity to their endogenous target, p53. Examination of available inhibitor-bound structures of MDMX reveal a conserved pharmacophore but the structures adopt distinct conformations away from the binding site. This implies that global motions of the protein might contribute to molecular recognition. The conformational heterogeneity in MDMX was further confirmed by collecting residual dipolar couplings (RDCs). Further investigations on both MDMX and MDM2 are necessary to uncover whether the flexibility of MDMX contributes to the differential binding to inhibitors.

viii

Finally, NMR relaxation methods and state-of-the-art high-power Carr-Purcell-Meiboom Gill (CPMG) relaxation dispersion measurements, the first documented application on an enzyme, were used to characterize the solution dynamics of GMPK and the changes in dynamics upon GMP binding. Substrate binding resulted in restricting the amplitudes of motion for backbone amide bonds within the picosecond-nanosecond timescale. Meanwhile, CPMG showed dispersion in both in the absence and presence of GMP, such that substrate binding did not quench dynamics within the microsecond-millisecond timescale. Interestingly, more residues are observed to have dispersion in the bound form, some near the C-terminal of helix 3, which has previously been proposed to be involved in product release. Current studies show that substrate binding affect different timescales of protein motion. Future work shall follow how motions within different timescales are affected as GMPK processes its substrates – such as, for instance, binding of ATP analogs within the ATP binding site or simultaneous occupancy of both substrate binding pockets. This paves the way for a complete picture of the relationship of function and dynamics in the conformational enzymatic cycle of a bi-substrate enzyme using GMPK as a model.

The current work illustrates some of the diverse applications of NMR on three unique systems that are also drug targets. Information collected here can be leveraged on future structure and dynamics studies as well as drug discovery efforts targeting any of these proteins.

ix

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT	viii
LIST OF TABLES	xiii
LIST OF FIGURES	xiv

CHAPTER 01: BASIC THEORETICAL CONCEPTS IN	
NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY 1	
CHAPTER 02: PROTEINS AT THE ATOMIC LEVEL 41	1
CHAPTER 03: STRUCTURE, DYNAMICS, AND	
DRUG DISCOVERY BY SOLUTION NMR 88	3
CHAPTER 04: GENERAL PROTEIN BIOCHEMISTRY	
METHODS AND RELEVANT ASSAYS13	36
CHAPTER 05: CONFORMATIONAL HETEROGENEITY OF THE	
p53-BINDING DOMAIN OF ONCOGENIC PROTEIN, MDMX	30

CHAPTER 06: IDENTIFYING SMALL MOLECULE INHIBITORS
AGAINST ONCOGENIC PHOSPHATASE, PRL3
CHAPTER 07: SOLUTION DYNAMICS OF HUMAN
GUANYLATE KINASE, GMPK
CHAPTER 08: NMR: CONCLUDING REMARKS AND
A FEW MORE INTERESTING APPLICATIONS
REFERENCES
CURRICULUM VITAE

LIST OF TABLES

Table	Page
1. A list of commonly used spins relevant to biomolecular NMR	9
2. Biomolecular structure statistics from the protein data bank	45
3. Dynamics toolkit of structural biology	78
4. Model-free models used	100
5. Common reagents used for stable isotope labeling in NMR	138
6. Some <i>E. coli</i> strains used for expression of recombinant proteins.	140
7. Relevant properties of PRL3, GMPK, and MDMX for purification	153
8. Candesartan derivatives from <i>in silico</i> screening	238
9. Specificity of olsalazine by phosphatase profiling	251

LIST OF FIGURES

Figure Page	e
1.1 UofL BCC magnets	
1.2 Visualizing the behavior of spins in response to a magnetic field7	
1.3 An illustration of spin precession13	3
1.4 Net magnetization and tipping15	5
1.5 Relaxation of spins to equilibrium18	3
1.6 Fourier transform of a sine function21	1
1.7 Scalar coupling between nuclei26	3
1.8 Dipolar coupling between nuclei29	9
1.9 A one-pulse experiment23	3
1.10 The HSQC pulse sequence	3
2.1 Structure of myoglobin43	3
2.2 The central dogma of molecular biology47	7
2.3 The twenty-one essential amino acids57	1
2.4 Hierarchy of protein structures53	3
2.5 IDP ensembles55	5
2.6 The crystallography workflow59	9
2.7 Protein cryo-microscopy63	3

2.8 Multi-temperature, multi-conformer crystallography	65
2.9 Blindmen and the elephant	68
2.10 AlphaFold2 structure prediction	71
2.11 Conformational landscape of a protein	75
2.12 Explicit MD simulation box	79
2.13 Distance distributions from DEER	82
2.14 smFRET monitoring DNA bending and nuclease activity	85
3.1 Solution NMR structure	91
3.2 NMR accessible timescales	94
3.3 Overview of the CEST experiment	102
3.4 Overview of the CPMG experiment	106
3.5 Residual dipolar coupling in partially aligned media	111
3.6 Extracting couplings from IPAP-type experiments	114
3.7 Schematic for ¹ H, ¹⁵ N-HSQC	118
3.8 Exchange regimes in NMR titration	126
3.9 Principle of fragment-based drug discovery	129
3.10 Examples of WaterLOGSY and STD-NMR	133
4.1 pET28a vector map	141
4.2 Protein expression protocols	145
4.3 Basic chromatography purification protocol	148
4.4 Ni-affinity and SEC profiles for PRL3	151
4.5 Schematic for 3D-NMR	155
4.6 Part of the HNCA of GMPK	157

4.7 ¹ H, ¹⁵ N-TROSY for apo PRL3 with backbone assignments	.160
4.8 ¹ H, ¹⁵ N-HSQC for MDMX-nutlin3a complex with	
backbone assignments	.162
4.9 ¹ H, ¹⁵ N-HSQC for GMPK with backbone assignments	.164
4.10 Sample PRL3 kinetics assay	.168
4.11 Olsalazine binding to PRL3 by ITC	.173
4.12 GMP binding to GMPK by ITC	.175
4.13 Nutlin-3a binding to MDMX by ITC	.177
5.1 Regulation of cellular activities by phosphorylation	.182
5.2 PRL3 expression correlates with poor patient prognosis	.189
5.3 PRL3 sequence and some notable features	.193
5.4 The catalytic mechanism of PRL3	.195
5.5 Structure of PRL3	.199
5.6 Some available small molecule inhibitors of PRL3	.202
5.7 Characterizing PRL3 binding of compound 43 and pentamidine	.206
5.8 NMR on the interaction of PRL3 and thienopyridone	.209
5.9 NMR on the interaction of PRL3.C49A and thienopyridone	.211
5.10 Risk and reward of drug discovery strategies	.215
5.11 Broad PRL3 inhibitors from the FDA drug screen	.218
5.12 Salirasib and candesartan are non-toxic	.220
5.13 Inhibition of PRL3-induced HEK293T migration	.223
5.14 Inhibition of migration of colorectal cancer cells	.225
5.15 Salirasib and candesartan	.228

5.16 Lineweaver-Burke analysis of candesartan and salirasib	231
5.17 Molecular docking of salirasib and candesartan	234
5.18 Virtual docking of a small DrugBank library	240
5.19 Interaction of olsalazine and PRL3 by NMR	244
5.20 NMR titration olsalazine onto PRL3	246
5.21 Dose-dependent inhibition of PRL3 by olsalazine	248
5.22 Conformational changes upon open-to-closed transition	255
5.23 Molecular dynamics simulation of PRL3	258
5.24 Some PRL3 mutations away from the active site	262
5.25 Mutations do not affect structure of PRL3	264
5.26 Phosphatase activity of PRL3 mutants	266
6.1 MDMX binds the first transactivation domain of p53	273
6.2 p53 peptide binding to MDMX/MDM2	277
6.3 Nutlin was designed based on the p53 pharmacophore	280
6.4 Nutlin-3a binding to MDMX and MDM2 by ITC	286
6.5 NMR titration of nutlin-3a into MDMX	288
6.6 Binding affinity from NMR titration	290
6.7 Structures of MDMX reveal heterogeneity in distal sites	293
6.8 Principal component analysis of existing MDMX structures	296
6.9 MDMX motions contributing to largest variance in PCA	299
6.10 Representative data from IPAP-HSQC analysis of MDMX	303
6.11 Residual dipolar couplings in MDMX	305
6.12 Comparison of RMSD based on structure and RDCs	308

6.13 SJ298 and WK298 are the worst and best fit, respectively		
6.14 Preliminary high-power RD applied on MDMX		
7.1 De novo pyrimidine and de novo/salvage purine		
synthesis pathways320		
7.2 Reaction cycle of human guanylate kinase		
7.3 Cellular functions of GMPK		
7.4 Solution structure of GMPK		
7.5 Open and closed conformations of GMPK		
7.6 ¹ H, ¹⁵ N-HSQC of GMP-bound GMPK336		
7.7 NMR titration of GMP onto GMPK		
7.8 GMP binding residues and GMP-BD show highest CSP		
7.9 NMR relaxation parameters for GMPK		
7.10 Order parameter of GMPK and response to GMP binding		
7.11 Microsecond-millisecond motion detected in GMPK		
by high-power CMPG351		
7.12 Helix 3 in relation to the GMP-BD and CORE domains		
7.13 Proposed EPR experiments based on AF2 predictions		

CHAPTER 01

BASIC THEORETICAL CONCEPTS

IN NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

"I look forward with great expectations to the future evolution of this awesome and beautiful technique, which has given me so many years of joy and excitement in studies of the molecules of life."

Kurt Wütrich, 2002 Nobel Lecture

In 2002, the Nobel Prize in Chemistry was co-awarded to Kurt Wütrich "for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution." This award is but one of 13 Nobel Prizes (in Chemistry, Physics, and Medicine) awarded in the area of magnetic resonance (1). These recognitions highlight the impact that magnetic resonance has in various disciplines, specifically the implementation of nuclear magnetic resonance (NMR) spectroscopy in fields like structural biology, biophysics, drug discovery, and metabolomics among others. This chapter will introduce some basic theoretical concepts underlying NMR spectroscopy and briefly, illustrate how these concepts can be used to study molecules.

NMR spectroscopy is a powerful method used to study materials at the atomic level. It has been particularly useful for investigating the three-dimensional structures of biomolecules. As it is based upon the interaction of nuclear magnetic moments, or spins, with an external magnetic field, NMR observables prove to be highly information-rich probes for individual nuclei within the material, such as a protein, that is under investigation. The ability of NMR to probe structure in solution is a significant advantage over other protein structural biological techniques as it allows for the determination of native conformations of proteins. Moreover, NMR is uniquely powerful in its ability to interrogate protein dynamics at a wide range of timescales while maintaining atomic resolution in close to physiological conditions.

Strong Magnets, Spin States, and the 'Insensitivity' of NMR. Nuclear magnetic resonance spectroscopy relies on the detection of the interaction of nuclear spins with an external magnetic field. For biological NMR experiments, particularly involving the study of proteins, these magnetic fields are typically at least 11.7 Tesla (T). As of writing, the strongest magnetic field available for use in solution NMR is 28.2 T. Informally, a 28.2 T magnetic is referred to as a "1.2 GHz" referring to the frequency at which the protein nuclear spin resonates in a magnet of this strength. That said, the "600 MHz" housed at the Brown Cancer Center (**Figure 1.1**) is, in fact, a 14.1 T magnet. For comparison, the magnetic field strength of the Earth is merely 0.0000305 T.

Figure 1.1. UofL BCC magnets. Two of the three instruments located at the Molecular Imaging Research Center at the UofL Brown Cancer Center. On the left is the "600 MHz" NMR or 14.1 T Oxford magnet. In March 2018, it was upgraded with a Bruker AVANCE NEO console and a Bruker Prodigy CryoProbe. All NMR data presented in this work were collected with this instrument. On the right is the "800 MHz" or 18.8 T magnet. The magnets are usually branded with the corresponding "MHz" as can be seen in the 800 (the 600 label is on the other side, not shown in the photo).



When a protein sample in an NMR tube (**Figure 1.2A**) is placed inside an NMR spectrometer, nuclear spins within the sample align with the external magnetic field (**Figure 1.2B, C**). Spin, in this context, refers to a quantum mechanical property that is intrinsic to NMR active nuclei. This spin is also called the spin angular moment. Nuclear spin angular momentum is defined as the following, where *h* is Planck's constant and *l* is the spin angular momentum quantum number:

$$\left|\vec{I}\right| = \frac{h}{2\pi}\sqrt{I(I+1)}$$

NMR active nuclei are those that have non-zero spin angular momentum. Nuclei with an even number of protons and neutrons have zero spin and are, therefore, 'invisible' to NMR. Those that have an odd numbers of protons and neutrons have non-zero spins and are NMR active. A list of some common spins used for biomolecular NMR is provided in **Table 1**. These NMR active nuclei also possess a magnetic moment, μ , given by the following expression where γ is the gyromagnetic ratio, a quantity that is related to the sensitivity of a given nucleus in NMR:

$$\vec{\mu} = \gamma \bar{I}$$

Outside the static magnetic field, the magnetic moments will be randomly oriented (**Figure 1.2B**). Upon placing a sample in the spectrometer, these

magnetic moments will orient themselves along the external magnetic field, which is, by convention, along the z-axis (**Figure 1.2C**).

$$\mu_z = \frac{h}{2\pi} \gamma m$$

The magnetic quantum number *m* takes on values from +/ to -/ in integral steps such that there are (2I + I) angular momentum states for a given *I*. For a spin-1/2 nucleus, there are two spin states: spin-up or α corresponding to m = +1/2 and spin-down or β corresponding to m = -1/2. The spin-up state aligns with the magnetic field and corresponds to the low energy state, while the spin-down state aligns against the magnetic field and is the high energy state. The energy difference between these two states depends on the magnitude of the external magnetic field, B_0 , (**Figure 1.2D**) and can be expressed as:

$$\Delta E = \frac{\gamma h}{2\pi} B_0$$

Additionally, the population of the spin-up and spin-down states follow the Boltzmann distribution that is given by the following expression where k_B is the Boltzmann constant and *T* is the temperature in the Kelvin scale:

$$\frac{N_{\beta}}{N_{\alpha}} = e^{-\frac{\Delta E}{k_b T}}$$

Figure 1.2. Visualizing the behavior of spins in response to a magnetic field. Samples are placed in an NMR tube (A). Before the tube is positioned inside the magnet, the spins in the sample are randomly oriented (B). Inside the magnet, the spins react to the external magnetic field, B_0 and orient themselves either with or against the direction of the magnetic field, which is by convention along the *z*-axis (C). The energy difference between the low energy, α - or spin-up state, and the high energy, β - or spin-down state, is proportional to the strength of the external magnetic field (D).



	Natural	al Gyromagne Spin e (%) Ratio (MHz	Gyromagnetic
Nuclei	Abundance (%)		Ratio (MHz/T)
¹ H	99.98	1/2	52.48
² H	0.016	1	6.54
¹³ C	1.11	1/2	10.71
¹⁵ N	0.37	1/2	-4.32
¹⁹ F	100.00	1/2	40.08
³¹ P	100.00	1/2	17.24

Table 1. A list of commonly used spins relevant to biomolecular NMR.

For protons exposed to a 14.1 T external magnetic field, like proteins placed inside the 600 MHz magnet at the Brown Cancer Center, at room temperature, the difference in the population of the spin states is about one proton for every 10,000 protons, slightly favoring the low energy α state. With the ratio between the two states being very close to unity, the net magnetic moment is very small. The signal observed in NMR arises from the energy absorbed by the low energy spins that allow them to transition into the higher energy state, and the energy emitted by the spins during the spontaneous transition back to their equilibrium states. As such, the signal is proportional to the population difference between the states. Since this population difference is quite small under normal conditions, NMR is typically regarded as a relatively insensitive method. Typically, large amounts of samples, such as highly concentrated protein solutions, are necessary to increase the observed signal. Also, since the difference in energy is related to the magnetic field strength, development of stronger magnets has provided increased sensitivity to various NMR experiments.

Precession, Resonance, and Relaxation. Under the influence of an external magnetic field, magnetic moments precess (**Figure 1.3**) around B_0 at a frequency called the Larmor frequency:

$$v_L = \frac{\gamma B_0}{2\pi}$$

Figure 1.3. An illustration of spin precession. The motion of the spin is analogous to a spinning top.



This frequency is equivalent to the frequency with an energy corresponding to the difference in energy between the spin-up and spin-down states. This precession of magnetic moments is analogous to the motion of a spinning top.

The transition from the low energy state to the high energy state can be achieved by irradiation with radiofrequency, RF, waves orthogonal to the external magnetic field. The transition occurs when the nucleus is in resonance – that is, when the energy of the applied RF field, B_1 , of frequency v matches ΔE :

$$E = hv$$
$$v = \frac{\gamma B_1}{2\pi}$$

The frequency of this RF wave is also equivalent to the Larmor frequency of a given nuclei.

Prior to the application of an RF wave, the magnetic moments precess around the external magnetic field. The magnetic moments of the individual spins sum up to M_0 , the net macroscopic magnetization or bulk magnetization (**Figure 1.4A**). It is this net macroscopic magnetization that the NMR spectrometer, in fact, records. At equilibrium, the bulk magnetization is aligned along the magnetic field or the *z*-axis by convention. As a vector, the components along the x- and y-axes of the bulk magnetization are zero. Upon application of the RF wave, this bulk magnetization is rotated into the x-y plane, or the transverse plane, where it continues to precess about the *z*-axis (**Figure 1.4B**). The magnitude of the components of the bulk magnetization M_0 in the transverse plane can be written

as the following where β is the tilt angle from the z-axis and ω_o is the Larmor angular frequency:

$$M_x = M_0 \sin(\beta) \cos(\omega_0 t)$$
$$M_y = M_0 \sin(\beta) \sin(\omega_0 t)$$

In the NMR spectrometer, a coil is mounted with its axis aligned in the x-y plane. The precession of a magnetization vector is detected by this coil as the vector induces a detectable current. The strength of the detected signal based on the induced current depends on the x-component, M_x , of the bulk magnetization vector.

Figure 1.4. **Net magnetization and tipping.** Under the influence of an external magnetic field, the spins line up with or against the external magnetic field, with the lower energy spins (lined up) slightly more populated. This results in a net magnetization that is lined up with the magnetic field (A). When an RF pulse (B_1) is applied along the *x*-axis, the net magnetization, M_0 , is tipped away from the *z*-axis and onto the xy- or transverse plane, initially along the *y*-axis (B).



After the RF wave that tilts *M*₀ away from the z-axis and into the transverse plane is applied, net magnetization begins to return to the equilibrium distribution along the external magnetic field (**Figure 1.5A**, **B**). This process is called relaxation. During relaxation of the spin, the strength of the detected signal in the transverse plane decays over time and is recorded as the Free Induction Decay or FID (**Figure 1.5C**).

There are two relaxation processes that occur as the bulk magnetization returns to equilibrium (**Figure 1.5B**). As the spins relax, the net magnetization parallel to the external magnetic field begins to approach M_0 . That is, the z-component of the magnetization vector increases over time *t*:

$$M_z = M_0 \left(1 - e^{-\frac{t}{T_1}} \right)$$

This relaxation process is called longitudinal relaxation or T_1 relaxation, referring to the time it takes to recover the initial maximum bulk magnetization, M_0 .

At the same time, as the spins relax to their equilibrium positions, the transverse components of the bulk magnetization, M_x , begin to decay:

$$M_x = M_0 \ e^{-\frac{t}{T_2}}$$

This relaxation process is referred to as the transverse relaxation or T_2 relaxation.
Figure 1.5. **Relaxation of spins to equilibrium.** The spins precess along the *z*-axis as they continue to return to equilibrium position after the RF wave is applied. After the RF wave, spins lose coherence and the net magnetization in the transverse plane decreases (A). As this happens, the net magnetization along the *z*-axis begins to recover as illustrated by a vector along *z* increasing in magnitude. The increase along *z* happens as the net magnetization along the transverse plane decreases (B). The NMR detects the signal along the transverse plane and thus, appears as a free induction decay where intensity is lost over time (C).



y

Time

| X

M_{xy}

As mentioned, the NMR spectrometer records an FID during an NMR experiment. The NMR signal is collected in the time domain as an FID, which is the strength of the induced current due to the precession of the bulk magnetization over time. NMR data is typically analyzed in the frequency domain. This transformation is done through a mathematical procedure called the Fourier Transform (**Figure 1.6**), relating the time domain f(t) to the frequency domain, $F(\omega)$.

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$

An NMR experiment can then be thought of as the manipulation of the bulk magnetization through the application of pulses and delays to tease out properties of the nuclei within the materials under investigation. **Figure 1.6. Fourier transform of a sine function.** The plot of a sign function with a frequency of 5 Hz against time (left). The Fourier transform of this function shows a peak at the expected frequency (right).



Chemical Shifts. Briefly discussed above, the NMR signal ultimately depends on the precession of spins in an external magnetic field in response to an RF wave. *How, then, is NMR able to distinguish between unique nuclei in a sample?* For instance, in GMPK, one of the proteins that will be discussed in a subsequent chapter, there are more than 1500 hydrogen nuclei.

As mentioned, when a sample is placed in a magnetic field, spins align to the magnetic field. In practice, however, the magnetic field experienced by the nucleus may not be equal to B_0 . Under the influence of an external magnetic field, electrons induce a small magnetic field, an induced field, that opposes or augments the external magnetic field. As a result, the nucleus will be under the influence of an effective magnetic field, B_{eff} , that is dependent on a shielding factor, σ :

$$B_{eff} = B_0(1-\sigma)$$

The shielding factor is ultimately dependent on the chemical environment around the nucleus under observation. As such, non-equivalent nuclei will ultimately be subject to effective magnetic fields of slightly different strengths. In other words, these non-equivalent nuclei will have unique Larmor frequencies. This unique resonant frequency is called the chemical shift.

By definition, the chemical shift δ , expressed in parts per million or ppm, is defined as follows:

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \ x \ 10^6$$

The references used are typically tetramethylsilane (TMS) for samples in organic solvents or sodium trimethylsilylpropanesulfonate (DSS) for aqueous samples with proton chemical shifts that are, by convention, set to 0 ppm. Expressed in ppm, chemical shifts are independent of the magnetic field and allow for comparison of data collected at different fields.

The chemical shift is sensitive to the environment of the nuclei and is affected by several factors including paramagnetic contributions from nuclei with non-*s* orbitals, anisotropy of neighboring bonds, hydrogen bonding, and solvent effects. They are thus powerful probes for the perturbation of a system. This will be revisited in subsequent chapters as perturbation of chemical shifts can be used to study the interaction between biomolecules.

Given this, it becomes apparent how NMR is a powerful tool to study biomolecules at atomic resolution.

Dipolar and scalar couplings. Chemical shifts are caused by the interaction of the small magnetic field induced by electrons modulating the effective magnetic field by nuclear spins. In addition to this, spins also interact with neighboring spins and these spin interactions are, in part, what makes it possible to carry out a diverse set of NMR experiments. For example, in the Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) block used in experiments like the Heteronuclear Single Quantum Coherence (HSQC) discussed below, the magnetization transfers between the amide proton and amide nitrogen are possible through *J*-couplings.

24

Another such interaction are dipolar couplings which are relevant in structure determination in NMR discussed in subsequent chapters. Dipolar couplings also contribute to spin relaxation, along with chemical shift anisotropy.

J-couplings, also referred to as scalar couplings or spin-spin coupling, are spin interactions that occur through bonds. J-couplings arise when two nuclei are connected by chemical bonds. Typically, *J*-couplings can be measured between nuclei separated by one $({}^{1}J)$, two $({}^{2}J)$, or three $({}^{3}J)$ bonds but 'long range coupling' between nuclei separated by more than three bonds is also possible. Specifically, these interactions are a product of the interactions between nuclear spins and electrons localized in the bonds connecting them. In the discussion above, a spin-1/2 nuclei adopts two spin states depending on their alignment with the external magnetic field. These nuclear spin states affect the spin states of their valence electrons, which in turn, affect the spin states of the electrons of the bonded nucleus, which then affects the spin state of the bonded nucleus. A simplified illustration in one- and two-bond systems is shown in **Figure 1.7A**. J-couplings, then, provide information on bond connectivities, as well as dihedral angles, in determination of structures, especially of small molecules. For example, scalar couplings manifest as splittings in the NMR spectrum as illustrated in Figure 1.7B. Furthermore, these couplings allow for magnetization transfers between nuclei connected by at most three bonds which is leveraged in designing pulse sequences for NMR experiments.

25

Figure 1.7. Scalar coupling between nuclei. Scalar couplings arise from the interaction of nuclear spins and electron spins within the bonds linking the interacting nuclei. Electron moments (arrows) in the bond are anti-parallel according to the Pauli exclusion principle. The relative orientation of the nuclear moment and the adjacent electron moment determines the relative energy. As illustrated, when the moments are parallel, this results in a higher energy state relative to when the moments are anti-parallel (A). In an NMR spectrum, this manifests as a 'splitting.' Here, H_A and H_X are coupled via a three-bond scalar coupling (${}^{3}J_{AX}$). Thus, the signal arising from each (δ) is split by the exact value of the scalar coupling (B).







Meanwhile, dipolar couplings refer to the direct interaction between two spins *through space*, as opposed to scalar couplings, which occur indirectly through interactions with electrons in bonds. As illustrated in **Figure 1.8**, dipolar coupling between two nuclei in a magnetic field depend on the orientation of the internuclear vector with respect to the external magnetic field. Specifically, the strength of the coupling depends on the angle, θ , between the external magnetic field and the internuclear vector:

$$E \propto \left(3\cos^2\theta - \frac{1}{2}\right)$$

In isotropic solutions, where molecules tumble randomly in solution, the internuclear vector is also oriented randomly and dipolar couplings average out to zero. Under anisotropic environments such as in partially aligning media, residual dipolar couplings (RDCs) arise. RDCs will be further discussed in subsequent chapters. While dipolar couplings are averaged out by reorientation of the molecule in isotropic solutions, dipole-dipole interactions contribute to relaxation. In the nuclear Overhauser effect (NOE), transfer of nuclear spin polarization between protons that are within close proximity to each other occurs via dipolar coupling. The strength of this interaction depends strongly on the distance separating the two interacting spins. This information, NOE distance restraints, are important for NMR structure determination as discussed in subsequent chapters.

Figure 1.8. Dipolar coupling between nuclei. Dipolar coupling arises from a direct interaction between two nuclear moments. This interaction is dependent on the angle, θ , between the inter-nuclear vector, *r*, and the external magnetic field, B_0 .



A Simple NMR Experiment. To illustrate how these basic concepts are applied in practice, a very simple NMR experiment will be considered. The most basic NMR experiment consists of a single pulse followed by data acquisition (**Figure 1.9**).

When a sample is placed in the NMR spectrometer, nuclei within the sample are placed under the influence of an effective magnetic field B_{eff} . Initially, the bulk magnetization is aligned with this magnetic field (**Figure 1.4B**). A pulse is then applied for a specific amount of time. In practice, the pulse width is optimized to maximize the signal by ensuring that the bulk magnetization is tilted into the transverse plane. The optimized pulse width corresponds to a '90° pulse.' After the pulse is applied, the transverse magnetization begins to relax – the transverse magnetization decays (T_2 relaxation) and the equilibrium magnetization begins to grow (T_1 relaxation) (**Figure 1.5**). This is recorded as an FID during the acquisition time.

Figure 1.9. A one-pulse experiment. The simplest NMR experiment consists of a single pulse, in this case, a 90° pulse, followed by acquisition. This sequence can be repeated as many times as necessary to improve signal-to-noise. A recycle delay is placed before the pulse to ensure that all spins return to the equilibrium position before the start of another cycle. The 90° pulse here corresponds to the one illustrated in Figure 1.4, which illustrates what happens to the net magnetization.



The signal-to-noise ratio (S/N) can be improved by repeating this sequence as many times as necessary and adding together the subsequent measurements. The S/N increases with the square root of the number of scans, *ns*:

$$\frac{S}{N} \propto \sqrt{ns}$$

At the beginning of the sequence is a delay called the recycle delay time (**Figure 1.9**). This ensures that all spins are back to equilibrium prior to the next measurement. After the experiment, the time domain data is then Fourier Transformed to get the frequency domain data.

HSQC, a routine heteronuclear NMR experiment. Figure 1.9 illustrates the simplest NMR experiment consisting of a single pulse. In the study of biomolecules, more complex pulse sequences are used to manipulate spin systems and obtain information-rich NMR observables. This section focuses on the heteronuclear single quantum coherence (HSQC) experiment, the simplest and most routinely used two-dimensional NMR experiment for the study of biomolecules. In fact, the ¹H, ¹⁵N-HSQC spectrum is typically referred to as the fingerprint spectrum of proteins. This sequence will be briefly discussed for illustration, but a detailed analysis of the HSQC experiment is available elsewhere (2).

The HSQC pulse sequence can be broken up into four parts: preparation, evolution, mixing, and detection. Here, the HSQC experiment is analyzed using the product operator formalism, a shorthand way of analyzing pulse sequences (*3*). The preparation period transfers the magnetization from ¹H to ¹⁵N via insensitive nuclei enhanced by polarization transfer (INEPT). During INEPT, the spin polarization is transferred from the high gyromagnetic ratio ¹H nuclei, which has a larger Boltzmann population difference, to the low gyromagnetic ratio ¹⁵N nuclei, which has a lower Boltzmann population difference. For simplicity, the ¹H spin will be represented as *I* while the ¹⁵N by *S*. In the beginning of the experiment, both spins will not have any transverse component and are aligned with the magnetic field. A pulse about *x* will rotate the magnetization toward the transverse plane and depends on the flip angle, β .

$$I_{z} \xrightarrow{\beta I_{x}} I_{z} \cos \beta - I_{y} \sin \beta$$
$$I_{z} \xrightarrow{\frac{\pi}{2} I_{x}} - I_{y}$$

In the case of the HSQC pulse program, the first pulse is a 90° pulse ($\pi/2$ in the equation above) tipping the magnetization on *I* into the transverse plane (point a in **Figure 1.10**). While magnetization is in the transverse plane, the chemical shift evolves and coherence is transferred from in-phase to anti-phase magnetization due to heteronuclear coupling (${}^{1}J_{NH}$).

$$-I_y \xrightarrow{(2\pi J_{IS}IZSZ)} -I_y \cos(\pi J_{IS}\tau) + 2I_x S_z \sin(\pi J_{IS}\tau)$$
$$-I_y \xrightarrow{(2\pi J_{IS}IZSZ)} 2I_x S_z$$

Here, τ is set to $(2 \ ^{1}J_{NH})^{-1}$ resulting in a total conversion from in-phase to anti-phase magnetization on the proton. A 180° pulse on both the *S* and *I* spins is applied in the middle of this process resulting in refocusing of the chemical shifts. Finally, a 90° pulse is applied on both spins. Thus, at the end of the INEPT/preparation period, magnetization is transferred to the ¹⁵N (S) and all proton magnetization is along the *z*-axis (point b in **Figure 1.10A**).

$$2I_x S_z \xrightarrow{\frac{\pi}{2}(-I_y + S_x)} - 2I_z S_y$$

The second period is characterized by chemical shift evolution of the S-spin, ¹⁵N. At this point, all proton magnetization is still along the *z*-axis. This evolution is represented as follows, with Ω_s representing the chemical shift of S:

$$-2I_z S_y \xrightarrow{\Omega_S Szt_1} 2I_z S_y \cos(\Omega_S t_1) - 2I_z S_x \sin(\Omega_S t_1)$$

A 180° pulse on the *I*-spin is applied in the middle of this period to refocus coherence transfer due to the heteronuclear coupling. At the end of the evolution period of time t_1 (point c in **Figure 1.10**), antiphase magnetization of the nitrogen is present. Chemical shift evolution is followed by a mixing period that transfers the

magnetization back to the protons (*I*). Two 90° pulses are applied on both the *I*and *S*-spins for this to occur followed by coherence transfer back to in-phase magnetization.

$$I_{z}S_{y}\cos(\Omega_{S}t_{1}) \xrightarrow{90^{\circ}(I_{y}+S_{x})} 2I_{x}S_{z}\cos(\Omega_{S}t_{1}) \xrightarrow{(2J_{IS}I_{z}S_{z})} I_{x}\cos(\Omega_{S}t_{1})$$

At the end of this mixing step, magnetization is transferred back to the proton such that this mixing period is also referred to as an inverse INEPT step. The chemical shift evolution of the proton magnetization (I_x) is what is observed in the experiment and its evolution is modulated by Ω_s , the chemical shift of the attached nitrogen (point d in **Figure 1.10**). The other term (I_xS_x) is a double quantum coherence that is eliminated in a process that is not covered here.

The final component in the pulse sequence is a pulse train on 15N that *decouples* the ¹⁵N from the proton during detection by locking the nitrogen in the transverse plane. This pulse train, as well as the 180° pulse in the evolution period, are important and prevents the appearance of doublets in the proton and nitrogen dimensions, respectively.

Figure 1.10. The HSQC pulse sequence. A pulse sequence for a common HSQC experiment. Thick filled bars correspond to 180° or π pulses, while narrower unfilled bars are 90° or $\pi/2$ pulses. The delay τ is set to $(2 \ ^{1}J_{NH})^{-1}$, where $^{1}J_{NH}$ is equal to ~95 Hz, the scalar coupling between the amide proton and amide nitrogen.



This chapter provided a basic background to the origin of an 'NMR signal' and provides a practical example by way of a simple 1-dimensional, single pulse experiment and a discussion of the HSQC measurement. Other practical aspects of setting up an experiment and analyzing the data will not be covered here in detail. Subsequent chapters will focus on the application of NMR to answer various biological problems. In these chapters, certain aspects of the experiment set-up and analysis will be discussed as needed.

CHAPTER 02

PROTEINS AT THE ATOMIC LEVEL

"In determining the structures of only two proteins we have reached, not an end, but a beginning; we have merely sighted the shore of a vast continent, waiting to be explored."

John Kendrew, 1962 Nobel Lecture

Myoglobin holds the honor of being the first protein whose structure has been experimentally determined (**Figure 2.1**). This was made possible by John Kendrew and his team using a technique that is now known as X-ray crystallography (4). By the time he and Max Perutz were awarded the Nobel Prize in 1962 for their work on globular protein structures, only one other protein structure has been determined, that of hemoglobin by Perutz (5). These experimental structures have opened the doors to the molecular world of proteins and biomacromolecules. Sixty years since, the field of structural biology has advanced tremendously and more protein structures have been determined (**Table 2**), just as Kendrew predicted in 1962. These structures, or more accurately models, significantly advanced the study of proteins and their functions and have also aided in identification of modulators that can be developed into therapeutics. Still, there remains more of

the 'vast continent' still waiting to be explored. This chapter provides an overview of the structural biology of proteins and the importance of protein structures in understanding their function, as well as a survey of tools available to study these structures. **Figure 2.1. Structure of myoglobin.** A Goodsell-like (left) and tube (right) representation of the myoglobin structure determined by John Kendrew and his team. The highly helical structure is apparent in the tube representation. The bound heme is colored in magenta. This is the first ever protein structure to be determined. (PDB: 1MBN)



Biomolecule*	X-Ray	NMR	EM	Multiple**	Other**
Protein only	147127	11969	7542	186	104
Protein/Oligosaccharide	8687	31	1324	0	0
Protein/Nucleic Acid	7757	277	2405	3	0
Nucleic Acid only	2448	1412	62	0	0
Oligosaccharide only	11	6	0	1	4
Other	154	31	5	0	0

Table 2. Biomolecular structure statistics from the protein data bank.

* Current as of 15 June 2022

** "Multiple" refers to structures determined using a combination of techniques such as X-ray and NMR or theoretical modeling and NMR, among others. Meanwhile, "Other" refers to structures determined using methods besides X-ray crystallography, NMR spectroscopy, and EM. These methods include solution scattering and infrared spectroscopy.

Proteins, workforce of life. Life is possible because of the interaction of biomolecules (biological molecules) that have various roles and functions to sustain it. One of the major types of biomolecules are proteins – along with carbohydrates, nucleic acids, and lipids - which are the workhorses of the cell. Following the central dogma of molecular biology (Figure 2.2), DNA is first transcribed into RNA, which is then translated into a chain of amino acids, the building blocks of proteins. This series of amino acids, also known as the primary sequence of a protein, to a certain extent, determines the shape the protein will take. The degree of diversity in amino acid sequence allows proteins to adopt a wide range of three-dimensional structures and consequently allow them to carry out an array of functions that are essential to life. There are proteins that catalyze chemical reactions, assist in the transport of molecules, provide cellular structure and scaffolding, among others. Thus, to understand how a cell functions - both in the healthy and diseased states - requires an understanding of how proteins function.

Figure 2.2. The central dogma of molecular biology. Information stored in DNA is first transcribed into RNA which is subsequently translated to the protein sequence. The central dogma involves two of the major classes of biomolecules, nucleic acids (DNA and RNA) and proteins. Structures used are just for illustrative purpose. (PDB: 1FV7, 1UBQ)



Protein structure: linear polypeptide to three-dimensional structures. Protein structure can be described in four levels. As discussed above, the primary structure of a protein simply refers to the sequence of amino acid residues that constitute it. Amino acids are organic molecules that contain amino (-NH₃⁺) and carboxylate (-COO⁻) functional groups and a side chain (-R) that determines their identity. There are 21 such amino acids (21 different side chains), including the rare but essential selenocysteine, that are encoded by the genetic code (**Figure 2.3**). In the protein primary structure (**Figure 2.4A**), these amino acids are linked together by peptide bonds.

The secondary structure of proteins refers to recurring local structures assumed by a sequence of amino acid residues in a chain which are typically α -helices, β strands, and loops (**Figure 2.4B**). Secondary structures form due to local interactions in the backbone of adjacent amino acids. With the now available structural information, propensities of amino acid residues to form helices or strands can be used to predict the secondary structure of a given polypeptide chain using several tools available (6–10).

The configuration of proteins in three-dimensional space is described by their tertiary or quaternary structures. Secondary structure elements come together to form the tertiary structure (**Figure 2.4C**) that is usually determined by the interactions of the amino acid side chains. In some cases, this tertiary structure determines the functional form of a protein. In other cases, the functional form of the protein may be several polypeptide chains or proteins that form a multi-subunit complex (**Figure 2.4D**), the quaternary structure. For instance, myoglobin is

49

monomeric and does not have a quaternary structure, while hemoglobin is a tetramer.

It is worth noting, however, that not all proteins possess three-dimensional structures. Intrinsically disordered proteins (IDP) exist as dynamic ensembles (**Figure 2.5**) and, while they may assume transient secondary structures, are typically disordered – lack 'ordered structure' – in their native milieu (*11–15*). The current work focuses on globular proteins and the discussion will be biased towards these proteins. NMR, however, is perhaps the most powerful tool to study IDPs and this will be pointed out in subsequent sections and chapters whenever relevant. After all, IDPs are a part of Kendrew's 'vast continent.'

Figure 2.3. The twenty-one essential amino acids. The primary sequence of a protein consists of a linear arrangement of amino acids from the pool of twenty amino acids encoded by the human genome. The amino acids are grouped based on their side chain properties: electrically charged, polar uncharged, special, and hydrophobic. This image was originally created by the Dancojocari and licensed under CC-BY-2.0. Image was used as is, and not altered in any way.



Figure 2.4. Hierarchy of protein structures. The primary structure consists of a string of amino acids linked by polypeptide bonds (A). Local interactions in a series of amino acids lead to the formation of secondary structures. The helix (cyan), loop (orange), and strands/sheets (magenta) are secondary structures that appear in protein structures (B). Long range interactions of secondary structures form the tertiary structure of a protein, simply referred to as the 'protein structure.' Secondary elements in *B* are shown in color within the 3D structure of PRL1 (C). Finally, the quaternary structure consists of multiple polypeptides that form a functional complex. One monomer is shown in color within the PRL1 trimer (D).


Figure 2.5. IDP ensembles. Intrinsically disordered proteins do not have the 'traditional' structures that folded proteins have. Shown is an example of an IDP ensemble, that of Sic1. Individual states in the 11-member ensemble are colored along a blue-to-red continuum. (PED: 00001)



A brief survey of tools for structural biology. The goal of structural biology is to uncover the three-dimensional structures of biomolecules to gain an understanding of their biological functions. After the completion of the Human Genome Project, one of the next frontiers is arguably the determination of the functions of all known proteins (*16*, *17*). Critical to this end is the experimental determination of the structure of these proteins (*17*). To date, the most common experimental methods used to determine protein structures are X-ray crystallography, NMR spectroscopy, and more recently, the emerging and rapidly developing field cryogenic electron microscopy (Cryo-EM) (*18*).

The Protein Data Bank (PDB), under the management of the Research Collaboratory for Structural Bioinformatics (RCSB), is an archive for biological macromolecular structures (*18*, *19*). As of May 2022, the PDB is home to more than 190000 structures of biomolecules. Of all the deposited entries in the PDB (**Table 2**), more than 88% were determined by X-ray crystallography – that is more than 150000 structures determined since the very first X-ray structure was determined in 1957 (*4*, *20*). NMR accounts for about 7% of the deposited structures while CryoEM is at 4% but rapidly catching up.

X-ray crystallography, still the technique of choice. That structures determined by X-ray crystallography dominate the PDB is not surprising. X-ray crystallography is applicable to a wide range of protein families and methodological and technological advances mostly kept up with the demands of the field (*21*).

In this method (**Figure 2.6**), a protein structure is determined by analyzing the diffraction patterns of a crystallized protein sample exposed to X-ray beams. Electrons in the crystallized protein sample deflect incident X-rays leading to a specific diffraction pattern. Using computational analyses, this diffraction pattern is then converted into an electron density map onto which a model of the protein structure can be fit (*22, 23*). Unlike NMR, discussed below, there is virtually no size limitation for X-ray crystallography (*24, 25*). However, the requirement for obtaining protein crystals limit its applicability to systems that can be crystallized or manipulated to be crystallized. Some systems that have eluded structure determination by X-ray crystallography include some membrane proteins and proteins that have highly flexible regions.

X-ray crystallography as a field has not remained static. Crystallographers continue to address challenging systems such as membrane proteins, through improvements in protein expression and solubilization techniques (*21*, *26*, *27*).

NMR: from spins to structures. Some basic principles and concepts concerning nuclear magnetic resonance were described in the previous chapter. As discussed, NMR spectroscopy relies on the interaction of nuclear spins with an external magnetic field. Experiments that are of pertinence to protein structure and dynamics will be further discussed in subsequent chapters.

Figure 2.6. The crystallography workflow. Protein samples are crystallized and subjected to beams of X-ray generating a diffraction pattern. Several reflections are obtained as the crystal is rotated to capture several different orientations. An electron density map is then derived from the diffraction data. Computational methods, and any other available data, are used to derive an atomic model. This image was originally created by Thomas Splettstoesser and licensed under CC-BY-SA-3.0. Image was used as is, and not altered in any way.



Several NMR observables can be recorded for a protein and can be used to analyze its conformation. For instance, chemical shifts are very sensitive probes of molecular structure (*28*). Protein structure determination by NMR begins with the assignment of observed resonances and the collection of restraints such as nuclear Overhauser effect (NOE) distance restraints which are used to inform structure calculations (*29–31*). These experiments are acquired in solution, close to physiological conditions. As such, structure determination by NMR does not require protein crystals, unlike X-ray crystallography; this is arguably one of NMR's biggest advantages over crystallography. While developments in hardware and software have significantly expanded the size limitations of NMR, *de novo* structure predication of relatively larger proteins, about 35 kDa or higher, remains challenging (*32–34*).

As will be further discussed later, however, the true power of NMR is realized in its ability to study protein dynamics (*35–38*).

CryoEM: The *cool* new kid in the block. Cryogenic electron microscopy is an emerging method that is beginning to parallel the resolutions obtained by X-ray crystallography, a long way from the era of *'blobology'* (Figure 2.7A) (39, 40). While majority of the structures determined by CryoEM have been relatively lower in resolution compared to X-ray crystallography, recent developments have enabled determination of structures at atomic resolution (41–43).

In CryoEM, protein samples are preserved in their hydrated states, frozen and embedded in ice (**Figure 2.7B**). If the freezing process is done quickly enough,

several protein states may be preserved allowing for structure determination of more than one conformation (44). The ability of CryoEM to visualize multiple protein conformations without the need for crystallizations is a significant advantage over X-ray crystallography. That said, multi-conformer X-ray crystallography (MMX) is also an emerging approach to study fluctuations in protein structures (**Figure 2.8**) (45). Like NMR, CryoEM has some size limitations as it is currently most useful for larger proteins and it has faced challenges in determining structures of sub-50 kDa proteins (46–48).

Figure 2.7. Protein cryo-microscopy. The evolution of CryoEM from the era of "blobology" to now reaching close to atomic resolution is artistically illustrated using the structure of galactosidase. This image was originally created by Veronica Falconieri and Sriram Subramaniam (NCI) and licensed under CC-BY-NC-2.0. Image was used as is, and not altered in any way (A). The process of structure determination by CryoEM begins by preserving the hydrated protein states frozen in ice. The Cryo-Transmission Electron Microscopy (Cryo-TEM) of GroEL in amorphous ice shows several particles of GroEL in different orientations. These different orientations are classified and used to reconstruct the three-dimensional structure of GroEL. This image was originally created by Vossman and licensed under CC-BY-SA-4.0. Image was cropped to better show the particles.



Figure 2.8. Multi-temperature, multi-conformer crystallography. The structure of the WPD loop of phosphatase, PTP1B collected at different temperatures. The electron density is fitted to a major conformation and a secondary conformation. As the temperature is increased, the population is shifted to the secondary conformation identified at 100 K (indicated with an arrow). This image was adopted from Figure 2C of *Keedy, et al. (2018) eLife 7:e36307,* licensed under CC-BY-4.0. Image was cropped and panels rearranged without further modification.



Integrating the tools of structural biology. In an Indian parable, a group of blind men who have never encountered or seen an elephant begin to come up with their idea of what the elephant looked like by simply touching one, and only one, part of the elephant (Figure 2.9). As expected, their individual descriptions of an elephant were largely different – a tusk and a tail might be described as long, but, of course, a tusk is not a tail and that is about where the similarities end. This parable can be used as a metaphor for the need for integrative approaches in structural biology. The techniques mentioned above as well as many others have their own strengths, and the integration of these data will enable tackling questions that might not have been considered tractable by a single technique (49-52). For instance, the structure of a dodecameric aminopeptidase TET2 was determined at high resolution by combining NMR and CryoEM (53). At the time, this combined approach to studying the half-megadalton complex exceeded the limitations of the individual methods – the size limitation for NMR and limited resolution for CryoEM. Computational methodologies, like structure prediction discussed below and simulation methods, also play a bigger role in integrating the results from various techniques (52, 54). There are several examples for successful integration of structural data and the future of structural biology is already shaping to be an integrative one (52, 54).

Figure 2.9. Blindmen and the elephant. Each blind man depicted can only examine one part of the elephant. They are then asked to describe an elephant, which they have never seen or encountered before, using only the data they collected. This image is public domain.



AlphaFold2 revolutionizes structure prediction. The 'Holy Grail' of structural biology is the protein folding problem: *how does the primary sequence of a protein dictate its tertiary structure* (55)? While DeepMind's AlphaFold2 (AF2) does not, technically, address this question, it has significantly advanced our capabilities to predict the 3D structures of proteins based on their primary sequences. In the 2020 Critical Assessment of Structure Prediction (CASP14), AF2 achieved a Global Distance Test median score of 92.4/100 which translates to an average error of about 1.6 Angstroms (56, 57). Subsequently, DeepMind applied AF2 to predict the structure of 98.5% of the human proteome and several other proteomes (56, 58).

The availability of high confidence predictions (**Figure 2.10**) has immediately impacted the field as a whole (*59*). In crystallography, AF2 predictions may be used to address the 'phase problem' accelerating data analysis (*60*, *61*). In calculating the electron density from diffraction patterns, the amplitude and phase of a wave corresponding to a diffraction spot is necessary. However, only the amplitudes can be determined experimentally. A known structure of a protein with strong structural similarity can be used to address this problem in an approach known as molecular replacement. AF2 predictions can now be used in this approach as has been shown in the case of the structure determination of Nmd4, a protein involved in nonsense-mediated mRNA decay (*61*). Similar applications in conjunction with other methodologies can easily be envisioned (*50*). Furthermore, the structural biology community has summarized several viable applications of AF2 models (*62*). Clearly, AF2 predictions have had tremendous impacts in the field and will continue to do so in coming years, particularly in integrated structural biology.

Figure 2.10. AlphaFold2 structure prediction. AlphaFold2 had tremendous success during the 2020 Critical Assessment of Structure Predictions (CASP14). Predications on most of the proteins encoded by the human genome are now housed in the AlphaFold Protein Structure Database. These predictions can have varying confidence levels. The AF2 model (gray) and recent MDMX N-terminal domain structure show good agreement, although arguably because several existing structures of this part of MDMX are part of the training set (A). The AF2 model (gray) captures some features of the 7TM domain of SSTR2, but with significant differences in the orientation of some of the helices (B). The full-length MDMX protein consists of a lot of intrinsically disordered regions, and AlphaFold2 low confidence regions (blue) correspond to these regions in MDMX (C).



Proteins have *structures* **not** *a* **structure.** So far, this chapter has focused on 'snapshots' of proteins and on several occasions, proteins have been referred to as having 'structure.' Meanwhile, in the brief discussion of IDPs (**Figure 2.5**) and in highlighting the capability of CryoEM to capture multiple conformations (**Figure 2.7B**), it is implied that proteins exist as ensembles of structures. The idea that proteins move in solution is far from novel. As early as the 1960's, observations that proteins assumed different conformations and that they oscillate 'between these conformations' have been made (63). Conventional structure determination can reveal snapshots of the busy life of a protein but to truly understand their functions, it is necessary to consider the breadth of conformation sampled by proteins, as well as the interconversion between these conformations.

The classical structure-function paradigm – 'structure determines function' – arising from early work by Anfinsen, thus, offers an incomplete picture of the 'life of proteins' (64, 65). Instead, a characterization of function necessitates a complete description of the protein's conformational energy landscape (**Figure 2.11**) (54, 66-68). The conformational landscape of a protein consists of an ensemble of protein structures – or conformational substates – that interconvert. In a 1991 paper, Frauenfelder and company described the motions of myoglobin in the context of its conformational energy landscape as it relates to binding to oxygen and carbon dioxide and also recognized that some motions may not be coupled to function (69). In a representation of the conformational energy landscape, analogous to the protein folding funnel diagram, the different structures a protein can adopt are in energy minima (70-73). Perturbations such as

mutations, post-translational modifications, or ligand binding, reshape this conformational energy landscape and redistribute the protein structures (**Figure 2.11**). Large-scale conformational changes are separated by larger energy barriers while smaller amplitude motions, like side chain rotations, are separated by smaller energy barriers. As alluded to by Frauenfelder, not all of the structures are functionally relevant (*69*). Perturbations that reshape the conformational energy landscape may enrich or decrease the population of functional states. In this regard, it is easy to see how conformational landscapes are linked to protein functionality. Arguably, then, the conformational energy landscape of a protein, rather than a single structure, determines its function.

Figure 2.11. Conformational energy landscape of a protein. A theoretical protein consists of four domains adopting four major conformations, with only one of the conformations A* capable of binding to a target protein. In wild-type or normal conditions (top), the functional A* is only a minor conformation. When perturbed (bottom), such as by an activating mutation or an activating post-translational modification, A* becomes the primary conformation. The conformational energy landscape is redistributed by a perturbation. Under normal conditions (top), states B and C, can easily interconvert and only require small amplitude domain motions. Meanwhile, the barrier to states A and D are much larger and larger amplitude conformational changes are required. In the perturb state, it is also possible to alter the relative energies between these conformations.



The dynamics toolbox of structural biology. The next chapter focuses on the applications of NMR spectroscopy to study protein structure and dynamics, as well as in drug discovery. As already mentioned, the true power of NMR is in its unique ability to interrogate dynamics of proteins (35–37). However, NMR does not have a monopoly on protein dynamics and some of these methods (though not exhaustive) are very briefly mentioned and described here. Some strengths and limitations are also summarized in **Table 3**.

Advances in theory and software have significantly improved the ability of simulations to provide insight related to the dynamics of biomolecules (74). With the ever-increasing computational capacity available, long simulations of really large systems are now possible allowing molecular dynamics (MD) simulation (**Figure 2.12**) to provide insight on the dynamics of proteins at increasing temporal and spatial scales(74–76). MD simulations can be used to explore a protein's full conformational space, the equilibrium conformational energy landscape of a protein, and the motions and dynamics of proteins. The accuracy of these simulations depends, of course, on the force field use, and as simulations, on the user's input and set-up, but these potential pitfalls are also an advantage as the system is fully manipulable by the user (77–*81*).

Method	Advantages	Limitations
NMR	Covers all timescales of motion; atomic resolution; near physiological; no chemical modification necessary	Size limitations
		(conventionally <~35 kDa);
		requires a lot of material;
		costly labeling and
		instrumentation
CryoEM	Multiple structures/conformations accessible; no crystallization required	Size limitations (applicable to
		bigger proteins); low
		resolution (but fast
		improving)
EPR	Virtually no size limit; large	Requires chemical
	distance range; equilibrium	modification; single site at a
	ensemble accessible by DEER	time (or distance)
FRET	Amenable to ensemble/single-	Requires chemical
	molecule analysis; low sample	modification; distance
	requirements; detects rare events	limitations
MD	Atomic level; long timescales (multiple microseconds accessible); user manipulable	Accuracy depends on force
		fields; still a simulation
		(although becoming
		accurate!)
Integrative	All timescales accessible; lots of	None, unless <i>you</i> don't like
Approach	information; all above strengths	collaborating!

Table 3. Dynamics toolkit of structural biology.

Figure 2.12. Explicit MD simulation box. Improvements in software and algorithms, including dedicated machines for molecular dynamics simulations, have made microsecond-long simulations accessible. Shown is an example of a simulation box used for the simulation of PRL3 in the current work consisting of explicit water molecules (sticks) and counterions (blue spheres). Simulations were done on the University of Louisville Cardinal Cluster. This box contains about 60,000 atoms. With Anton 2, a day of simulation can cover upwards of 50 µs for this system size.



X-ray crystallography described above also provides some dynamics information by way of the temperature factors associated with experiments (*82*, *83*). These B-factors provide a hint on potential flexibility of proteins and B-factors obtained at room temperature, indeed, correlate well with solution dynamics (*82*). Furthermore, algorithms have been developed that are able to fit and recognize multiple conformers in a single electron density map (*84*). This provides insight on the conformational equilibrium of the system under investigation. Multi-temperature and multi-conformer crystallography (MMX) capture conformational changes as a function of temperature (**Figure 2.8**) and a similar approach can be applied to other perturbations such as pressure, pH, ligand concentration, and other perturbations that can be applied in series (*45, 84, 85*).

Electron paramagnetic resonance (EPR) spectroscopy combined with sitedirected spin labeling of proteins provide a way to investigate protein dynamics in diverse molecular systems with no size limit (*86*, *87*). EPR is able to easily monitor spin label dynamics and solvent accessibility that provide information on side chain motion and secondary structures. Meanwhile, pulsed electron double resonance (PELDOR), also known as double electron electron resonance (DEER), permits the determination of the distance between a pair of spin labels (**Figure 2.13A**) (*88– 91*). PELDOR/DEER ultimately provide a histogram of distances reflecting the ensemble nature of proteins. Perturbations that alter the conformational energy landscape of a protein are reflected as changes in the characteristic of the histogram and the relative populations and distributions of detected states (**Figure 2.13B**).

Figure 2.13. Distance distributions from DEER. PELDOR/DEER requires the incorporation of spin labels (represented here as red spheres). The distance between a pair of spins is then measured. In this hypothetical protein scenario, perturbing Condition A and putting the protein in Condition B increases the separation between the labelled helices by almost 2 nanometers (A). This is reflected in the shift in plot of P(r) vs. distance towards longer distances (B). It is also possible to have both states highly populated in each condition and are in equilibrium with each other (not shown). (PDB: 3TT1, 3TT3)



Förster resonance energy transfer (FRET) is a method that can interrogate protein dynamics at the ensemble or single molecule (smFRET) level. It relies on non-radiative energy transfer between two fluorescent probes – a donor and an acceptor – reporting on their intervening distance (**Figure 2.14**) (*92*). smFRET is particularly powerful in resolving transition between states and unlocking kinetics information that is not accessible in ensemble-based experiments such as PELDOR/DEER (*93–95*).

In the era of integrative structural biology, these aforementioned methods can be used to complement each other as well as other techniques including small angle X-ray scattering, hydrogen-deuterium exchange mass spectrometry, and NMR (discussed in its own chapter), among others to fully characterize the dynamics of biomolecules and how they relate to their structures and functions (*49*, *50*, *54*, *96*). **Figure 2.14. smFRET monitoring DNA bending and nuclease activity.** FRET relies on the energy transfer (E) between donor (D) and acceptor (A) fluorophores. In single molecule FRET, transition between states is resolved and not ensemble averaged. This permits for the monitoring of conformational changes as perturbations happen. Here, a flap endonuclease (FEN1), initially bends the substrate, non-equilibrated (NonEQ) DF-6,1_{flap}, prior to cleavage. Each step is clearly resolved in the time series. The reaction was monitored using an Alexa Fluor-647 acceptor (red star) and a Cy3 donor (green star). This image was adopted from Figure 2A of *Rashid, et al. (2017) eLife 6:e21884,* licensed under CC-BY-4.0. Image was cropped without further modification.



This chapter presented a brief overview of proteins and the structural biology toolkit to study the structure and dynamics of proteins. All of these tools have their advantages and limitations, and each contribute a view of protein structures and their motions. As the 'low-hanging fruits' of structural biology are addressed, advances in these tools and ways to integrate them will usher in significant advances in the understanding of the relationships between protein structure, function, and dynamics.

CHAPTER 03

STRUCTURE, DYNAMICS, AND DRUG DISCOVERY BY SOLUTION NMR

"The complementarity of NMR to other structural techniques is such that as they continue to advance so too will the utility of NMR. If anything, NMR is far more valuable today than it was even a decade ago."

> Lewis Kay The FASEB Journal 2018 (97)

The ability of NMR to study proteins in solution at atomic resolution and without any need for modification definitely gives it an edge over other techniques, especially in the dynamics of proteins. In addition to conventional structure determination, NMR is particularly useful in probing dynamics across a wide range of temporal scales, from picoseconds to hours and beyond (*36*, *98*, *99*). In this chapter, methods that are used in this body of work are discussed, as well as some other applications that are related to drug discovery, structure determination, and the study of protein dynamics. Note that most discussions will be limited to the applications rather than the theory and practical aspects. **Solution structures of proteins.** Structure determination by NMR relies on the collection of distance and orientational restraints and the implementation of these restraints in structure calculation (*30*, *100–102*). In structure calculation software like CYANA, the structure calculation method is implemented as a minimization problem using a target function and the experimental constraints, starting from a polypeptide chain (*30*, *102*). Thus, NMR structure determination requires that as many chemical shifts are assigned as possible.

Structure determination typically starts with the sequential assignment of all backbone resonances, as well as assignment of side chain resonances. These are usually accomplished using a set of routinely used 2D and 3D NMR experiments (*103–106*). These set of multi-dimensional experiments use correlations among resonances, such as that between the amide proton and amide nitrogen in the HSQC experiment discussed in Chapter 1, that aid in sequential and side chain assignments. An example of a 3D experiment that is most routinely used for backbone assignment is the HNCA which uses correlations from the amide proton, amide nitrogen, and alpha carbons in the backbone. This will be discussed further in the next chapter, particularly applied to the systems studied in the current work.

The distance restraints used in NMR structure calculations are also referred to as NOE distance restraints, for 'nuclear Overhauser effect' (*102*, *107*). The NOE arises from the interaction of two spins that are within close proximity such that transfer of nuclear spin polarization from one spin to another can occur via cross relaxation (*108*, *109*). The dependence on distance is a consequence of the fact that dipole-dipole interactions are distance dependent. The intensities of NOE
cross-peaks are related to the distances between the spins. While the intensities do not translate to an exact distance between the interacting spins, they can be converted to a distance interval that are as restraints in structure calculations (*107*). Observable NOEs are typically confined to about 5 Angstroms of distance between two interacting spins.

Meanwhile, orientational restraints can be obtained from residual dipolar couplings (RDCs), discussed more below. RDCs provide long-range orientational information as opposed to the distance-limited NOEs (*110*, *111*).

Other restraints typically used in NMR structure determination are dihedral angle restraints obtained from *J* couplings, chemical shifts which provide insights on the secondary structure, and hydrogen bond restraints (*28*, *30*, *101*, *105*, *112*).

All available information is used in an iterative structure calculation process that relaxes a random structure until all available experimental restraints are satisfied. As the restraints are provided as intervals – NOE and *J* couplings, for instance, do not provide exact distances and torsion angles – the final calculations yield multiple structures that satisfy the experimental restraints. Typically, the twenty structures with the lowest energy after minimization, all equally valid solutions, are selected (**Figures 2.5 and 3.1**) (*30, 112*).

Figure 3.1. Solution NMR structure. The solution structure of human guanylate kinase in its open form. The individual, equally valid members of the structural bundle are shown. NMR restraints, such as the NOE, are employed as an interval or an allowable range, resulting in the minimization yielding multiple solutions. This also allows NMR to be suitable in the study of IDPs as shown in **Figure 2.5**. (PDB: 6NUI)



The collection of a large number of datasets can make structure determination by NMR tedious but the lack of need for protein crystals makes it a powerful tool for proteins that are not amenable to crystallization. Automation of the structure determination process is an on-going area research, as well as the use of NMR observables in combination with other techniques (as alluded to by the introductory quote from Lewis Kay) make the process more accessible, even for larger proteins (29–31, 101).

NMR and accessible time window for protein dynamics. Protein dynamics occur at a wide range of temporal scales (*35*, *36*, *38*, *113*, *114*). Atomic vibrations, for instance, occur on the order of femtoseconds, while some protein folding events can take up to minutes or even hours. Molecular tumbling can occur in the picosecond-nanosecond timescale and collective motions typically occur in the picosecond to millisecond regime. Larger amplitude conformational changes, meanwhile, usually happen on the order of microseconds to milliseconds or longer (*114*).

As mentioned, NMR is a tool that is suitable to interrogating these vast temporal scales, and thus to probe protein dynamics (**Figure 3.2**). Some of the methods used in the study of dynamics, focusing on those used in this body of work, will be discussed more in this chapter.

Figure 3.2. NMR accessible timescales. NMR is able to probe a wide range of timescales. Relaxation experiments (R_1 , R_2 , *hetNOE*) can detect fast motions up to the overall tumbling motion, τ_c . Conventional relaxation dispersion (RD) experiments can detect slower motions but are currently limited to motions no faster than ~40 µs. Meanwhile, residual dipolar couplings (RDC), cross-correlated relaxation rates (CCR), J-couplings, paramagnetic relaxation enhancement (PRE), and chemical shifts can quantify the amplitude of motions from fast to slow, including the gap from about to τ_c to approximately 40 µs, which has been referred to as the 'hidden' time window. However, these quantifications do not include the characterization of kinetics, such as conformational interconversion or molecular recognition occurring within this hidden time window. The high-power RD recently developed has now allowed characterization of motions as fast as ~3 µs, and even faster motions at ~400 ns under super-cooled conditions. Image by David Ban (unpublished).



NMR relaxation methods provide ideal tools for studying motions faster than rotational correlation time or tumbling time, τ_c , of a protein, which is approximately several nanoseconds (115). The tumbling time refers to the average time it takes for a protein to rotate in solution. This value is dependent on protein size and shape and the viscosity of the solution. The amplitudes of these motions are quantified with the Lipari-Szabo (S^2) order parameter (116). Such motion has been proposed to contribute mostly to the entropy of proteins. Slow time scale motion, from around 40 µs to several ms can be probed by relaxation dispersion (RD) measurements (117). Such RD experiments are sensitive to conformational changes of proteins. Slow time scale motions have also been detected in several cross-correlated relaxation experiments (118, 119). Motions slower than the millisecond time scale are usually not averaged out in the chemical shift scale and thus, exhibit distinct peaks. By following these peaks using techniques such as exchange spectroscopy (EXSY), chemical exchange saturation transfer (CEST), and real time experiments, dynamics can be also investigated (117, 120, 121).

A time range of approximately four orders of magnitude, from τ_c to 40 µs, was "hidden" due to the lack of tools required to make experimental observations within this time range in physiological environment (*117*, *122*, *123*).

Important methods have been developed to access dynamics occurring within the hidden time window: residual dipolar couplings (RDC), cross-correlated relaxation (CCR) rates, scalar (J) couplings, paramagnetic relaxation enhancement (PRE), and chemical shifts (*119*, *122*, *124–128*). Particularly, with the model-free analysis of RDCs, the hidden time dynamics of ubiquitin have been

probed with RDCs using multiple alignment media and were found to be important for molecular recognition process between proteins (*129–131*).

NMR relaxation and the order parameter. NMR relaxation methods such as longitudinal (R_1) and transverse (R_2) relaxation methods are observables that report on global and local dynamics that occur in the picosecond-nanosecond timescale, or fast motions in proteins (*115*). These two relaxation events were discussed in Chapter 1 (T_1 and T_2 , corresponding to longitudinal and transverse relaxation times, respectively). Briefly, R_1 relates to the recovery of the bulk magnetization, while R_2 relates to the decay of the transverse magnetization (**Figure 1.5B**). Heteronuclear NOEs (hetNOE), meanwhile, are the most sensitive among the NMR relaxation observables to fast internal motions. Another observable used in studying these internal dynamics is the transverse cross-correlated relaxation rate (η_{xy}) (*115*, *132–134*).

The aforementioned NMR relaxation parameters contain information that reports on the fast internal dynamics of bond vectors within a protein under study. Using the model-free formalism, these relaxation parameters can be analyzed and summarized as an order parameter, S^2 , which describes the spatial restriction of the motion of a bond (*116*). This order parameter ranges from 0 to 1, with 1 corresponding to perfectly rigid. Importantly, this order parameter reports on motion in the picosecond-nanosecond timescale (*135*, *136*).

A spectral density function is used to fit the relaxation data. In this study, five models are considered, and the relaxation parameters are expressed as linear

combinations of a spectral density function (*135*, *137*, *138*). Three versions of the spectral density functions are used and are as follows.

$$J_{1}(\omega) = \frac{2}{5} \frac{S^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}}$$
$$J_{2}(\omega) = \frac{2}{5} \left(\frac{S^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}} + \frac{(1 - S^{2})(\tau_{fast})}{1 + (\omega \tau_{fast})^{2}} \right)$$
$$J_{3}(\omega) = \frac{2}{5} \left(\frac{S_{f}^{2} S_{s}^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}} + \frac{S_{f}^{2} (1 - S_{s}^{2}) \tau_{slow}}{1 + (\omega \tau_{slow})^{2}} \right)$$

The overall molecular tumbling time is τ_c . Meanwhile, τ_{fast} and τ_{slow} capture fast (τ_e , < 150 ps) and slow ($\tau_c > \tau_s > \tau_e$) internal motions.

$$\tau_{fast} = \frac{\tau_c \tau_e}{\tau_c + \tau_e}$$
$$\tau_{slow} = \frac{\tau_c \tau_s}{\tau_c + \tau_s}$$

The NMR relaxation parameters, R₁, R₂, hetNOE, and η_{xy} are expressed as linear combinations of a spectral density, depending on the model. The models and corresponding parameters are summarized in **Table 4**. The best model was selected using the Akaike's Information Criterion (AIC) described below (*135*, *138*).

$$AIC = \sum_{i=1}^{n} \left[\frac{(R_i - R'_i)^2}{\sigma_i^2} \right] + 2k$$

Here, *R* represents the experimental data and *R*' the back-calculated data for data of size *n* at each residue *i*. The number of adjustable parameters is taken into consideration as *k*, while experimental error is σ .

The generalized order parameter, overall, describes the dynamics of a bond vector at the picosecond to nanosecond timescales using NMR relaxation parameters (*116*).

Table 4. Model-free models	used.
----------------------------	-------

Model No.	Parameters	Spectral Density Function
1	$\{S^2, \tau_c\}$	$J_{1}(\omega) = \frac{2}{5} \frac{S^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}}$
2	$\{S^2,\tau_e,\tau_c\}$	$J_{2}(\omega) = \frac{2}{5} \left(\frac{S^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}} + \frac{(1 - S^{2})(\tau_{fast})}{1 + (\omega \tau_{fast})^{2}} \right)$
3	$\{S^2, R_{ex}, \tau_c \}$	$J_1(\omega) = \frac{2}{5} \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2}, R_2 = R_{2,0} + R_{ex}$
4	$\{S^2, \tau_e, R_{ex}, \tau_c \}$	$J_{2}(\omega) = \frac{2}{5} \left(\frac{S^{2} \tau_{c}}{1 + (\omega \tau_{c})^{2}} + \frac{(1 - S^{2})(\tau_{fast})}{1 + (\omega \tau_{fast})^{2}} \right),$ $R_{2} = R_{2,0} + R_{ex}$
5	$\{S^2, S^2_{f}, S^2_{s}, \tau_s, \tau_c \}$	$J_3(\omega) = \frac{2}{5} \left(\frac{S_f^2 S_s^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{S_f^2 (1 - S_s^2) \tau_{slow}}{1 + (\omega \tau_{slow})^2} \right)$

Some comments on the hidden time window. The 'hidden time' window is a range that covers approximately four orders of magnitude from the overall tumbling time (nanoseconds) to up to 40 μ s. This window is so-called 'hidden time' as is inaccessible by conventional experimental methods. In particular, the kinetics of motions occurring within this timescale has not been accessible (99, 117). Interestingly, motion occurring within this window has been shown to be critical for biological function. RDCs, discussed further below, are able to characterize the amplitude of motions occurring within the 'hidden time' window and has been used to show that such motions are involved in molecular recognition in ubiquitin (123, 139). While RDCs are able to capture the amplitude of motions from picoseconds to milliseconds including 'hidden time,' RDCs are not able to provide any information on the kinetics of motions.

Some advances (very briefly covered below) in relaxation dispersion experiments, however, have started pushing the limits of NMR well into the hidden time scale. That said, with the evidence that motions within this regime are important to function, there is no doubt that continued advances in methodology will eventually be able to probe the amplitude and kinetics of hidden time dynamics.

CEST and invisible states. When proteins undergo conformational exchange, they do so by adopting unique conformations that result in distinct magnetic environments for their spins (**Figure 3.3A, B**). This can be detected by various NMR experiments, including relaxation dispersion (discussed below) and chemical exchange saturation transfer (CEST) (*99, 117*).

Figure 3.3. Overview of the CEST experiment. Consider a system in chemical exchange where the population of the major conformation is >95% and $k_{ex} = k_a +$ k_b (A). The magnetic environment of spins in the major and minor conformations are distinct and would lead to unique peaks in the HSQC. The red peak corresponds to a typical signal observed in a ¹H,¹⁵N-HSQC, and corresponds to a spin in the major conformation. Meanwhile, the spin corresponding to the minor conformation, the light blue signal, is in a distinct position, but is not visible due to its low population. The CEST experiment works by applying a weak RF field, B_1 , at varying transmitter frequencies that sweep the entire ¹⁵N spectral width. The scenario when the fields are applied on resonance to the major peak and to the minor peak are illustrated with lightning and yellow dashed lines representing the transmitter frequency (B). The ratio of the intensities in the CEST experiment are normalized to a reference spectrum. When the field hits the peak, it modulates the signal leading to a dip in the CEST profile (C). The CEST profile is for the red residue in B. When the CEST field is applied where the invisible spin 'ought to be,' another minor dip is observed in this CEST profile due to saturation transfer between the exchanging major and minor states.



In a simple two-state exchange, a protein may adopt state A or state B, potentially resulting in a spin experiencing a unique magnetic environment for each state (**Figure 3.3B**). Originating in the 1960's, 'saturation transfer' experiments showed that perturbation of a spin in state A results in a transfer of the perturbation to the corresponding exchanging spin in state B due to chemical exchange (*140*).

In the study of protein dynamics, the most widely implemented version of CEST is the amide ¹⁵N CEST measurement. Here, a series of weak B_1 fields is applied at varying B_1 offsets covering the entire amide region of the system under investigation (**Figure 3.3B**) (141). These fields are applied for time, T_{EX} . Exchange is detected by plotting the intensities of the signals, I, to a reference intensity, I_0 , at T_{EX} = 0. A dip in this plot, which occurs when there is a loss of intensity (lower *I*) is observed when the weak-continuous-wave field is resonant with the exchanging spins (Figure 3.3C). CEST is able to detect 'invisible' or excited states – those that are very lowly populated and transiently formed (141). CEST is useful in detecting these invisible states and quantifying the kinetics of exchange that are within the millisecond-second timescale. In addition, CEST profiles also provide information on the longitudinal and transverse relaxation rates. It was recently shown that the R_1 and R_2 rates from CEST, parameters that report on the picosecond-nanosecond timescales as described above, are of sufficient accuracy and precision to be used to characterize protein dynamics within these timescales (142).

Overall, CEST is a powerful NMR method that is able to characterize slowly exchanging processes by providing not only exchange rates but also chemical shift information on invisible states. Furthermore, and as is used in this study, CEST

allows for the rapid determination of relaxation rates that characterize fast timescale backbone dynamics in a single measurement.

High power relaxation dispersion. Another NMR method used to probe chemical exchange is relaxation dispersion (RD). RD spectroscopy is able to capture motions occurring from about 40 μ s to milliseconds. These motions are slower than the internal fast dynamics captured by NMR relaxation methods and also slower than molecular tumbling but faster than the motions captured by CEST. RD spectroscopy provides information on the kinetics (k_{ex}), thermodynamics (population of states), and structure (chemical shifts) of a protein at atomic resolution (99, 117, 120).

Similar to CEST, RD experiments exploit the distinct conformational states of a system undergoing chemical exchange (**Figure 3.4A**). This chemical exchange between non-equivalent spin states generates another source of relaxation that contributes to the linewidth of resonances (*120*). This contribution, R_{ex} , is included in the effective transverse relaxation rate, $R_{2,eff}$, such that it is now enhanced compared to the intrinsic relaxation, $R_{2,0}$. The exchange contribution can be determined using Carr-Purcell-Meiboom-Gill (CPMG) or transverse rotating frame experiments (*143, 144*).

Figure 3.4. Overview of the CPMG experiment. Consider a system in chemical exchange where the population of the major conformation is >95% (A). The magnetic environment of spins in the major and minor conformations are distinct and the chemical shifts are unique, like the case in Figure 3.3 for the CEST experiment. In a CPMG RD measurement, the spacing between a series of refocusing (180°) pulses (black rectangles) are varied over a series of measurements where the relaxation time period is constant. Consider the extreme cases in (B) where the 180° pulses in (1) have more spacing between them than in (2). If there is chemical exchange occurring in between the pulses, the result is a loss in intensity (an increase in peak linewidth), and manifests as an increase in $R_{2,eff}$ due to the contribution from R_{ex} . When there is less time between pulses such as in (2) where there are more 180° pulses, the chances that exchange will occur in between pulses is lower, no loss in intensity is observed. The effect of the pulses in (B) correspond to the data point in (C) with the matching number. The red circles can be fit to an RD curve. If no exchange is happening, there will not be an RD curve. Accurate measurement of the exchange contribution requires that it be fully quenched, resulting in only $R_{2,0}$ contributing to the linewidth. High-power CPMG RD, where in the extreme case there is no spacing in between the 180° pulses, ensures this for systems that exchange in the lower microsecond to millisecond timescales.



In CPMG experiments, chemical exchange (**Figure 3.4A**) manifests as a dispersion curve in a plot of the dependence of $R_{2,eff}$ on the applied CPMG frequency. The CPMG experiment involves the application of a refocusing 180° pulse that is repeated with varying inter-pulse delay, τ_{cp} , that is related to the CPMG frequency.

$$\nu_{CPMG} = \frac{1}{\tau_{cp}}$$

This CPMG frequency manipulates the exchange contribution (**Figure 3.4B**). At higher frequencies, the contribution is dampened until it is completely quenched to $R_{2,0}$. Spins that do not undergo conformational exchange within the detectable timescales do not show any dispersion (**Figure 3.4C**) (*143*). The exchange contribution depends on the population weighted chemical shift variance, ϕ_{ex} , such that some residues that are involved in chemical exchange might not be detected if the chemical shift difference ($\Delta\omega$) between the two states is small.

$$\Phi_{EX} = p_A p_B \Delta \omega^2$$

The conventional CPMG can apply up to a 1.5 - 2.0 kHz CPMG (¹⁵N) frequency, although in the case of faster motions this is sometimes insufficient to quench the exchange contribution. This limits the detectable motions to those occurring no faster than ~40 µs (99, 117, 145). In this work, we employ high-power CPMG

extending the frequency limit to up to 6 kHz, covering up to 25 μ s of motion for the ¹⁵N nuclei (99, *117*, *145*).

After identifying spins that undergo chemical exchange based on dispersion in their CPMG profile, the spins are fit to various models to extract kinetic parameters (*146*). Under fast exchange, for instance, the dispersion profile can be fit to the Luz-Meiboom (LM) exchange model (*147*).

$$R_{2,eff} = R_{2,0} + \frac{\Phi_{EX}}{k_{ex}} \left(1 - \frac{4\nu_{CPMG}}{k_{ex}} \tanh\left(\frac{k_{ex}}{4\nu_{CPMG}}\right) \right)$$

In this model, R_{2A}^{0} and R_{2B}^{0} are the intrinsic relaxation rates for states A and B, respectively and k_{ex} is the exchange rate between states A and B (143, 148). Other models used to fit relaxation dispersion data include Bloch-McConnell and Carver-Richards (CR) (146, 149, 150). LM and CR are simplifications of the Bloch-McConnel equations which provides a general solution but is computationally expensive. CR is used to fit data in slow exchange.

Relaxation dispersion methods continue to be improved and its coverage extended into faster motions within the 'hidden' timescales. As discussed above, for ¹⁵N, this limit is currently at 25 μ s. However, using other nuclei, faster motions can be accessed such as ¹³C up to 10 μ s and ¹H up to 2.5 μ s (*122*, *145*). The use of nanoparticles has also been explored to modulate chemical exchange of a system to be detectable by RD (*151*, *152*).

It is worth noting that, in practice, a CPMG experiment is not routinely applied on systems without any indication of chemical exchange. Evidence for chemical exchange may include increased line broadening or data that fit into models that include an exchange, R_{ex} , term in the model-free analysis.

More on Residual Dipolar Couplings. Residual dipolar couplings (RDCs) were initially implemented as orientational restraints, as discussed in previous sections (*110*). RDCs arise from the dipolar coupling of two spins in a magnetic field. The bond vector connecting the interacting spins form an angle with the external magnetic field (**Figure 3.5A**). RDCs report on the orientation or direction of this bond vector (*110*, *123*).

$$D_{k}^{exp} = D_{ij}^{max} \langle (3\cos^{2}\theta_{k} - 1)/2 \rangle$$
$$D_{ij}^{max} = -\frac{\mu_{0}\gamma_{i}\gamma_{j}\hbar}{4\pi^{2}r_{ij}^{3}}$$

Here, μ_0 is the permeability of vacuum, γ_x is the gyromagnetic ratio of spin *X*, \hbar is the reduced Planck's constant, r_{ij} is the distance between nuclei *i* and *j*, and θ_k is the angle between the inter-nuclear vector and the magnetic field B_0 . For the N-H^N bond vector, the distance is assumed to be fixed at 1.02 Angstroms.

Figure 3.5. Residual dipolar coupling in partially aligned media. Dipolar couplings result from the interaction of two spins in a magnetic field. The vector connecting the two interacting spins, not necessarily bonded, forms an angle with the external magnetic field, which itself is defined to be along the *z*-axis (A). In an isotropic solution, these dipolar couplings are randomly oriented, and no residual dipolar coupling (RDC) is observed (B). In an alignment media (such as bicelles, phages, or in the presence of DNA), partial molecular alignment is induced and the RDC does not average out to zero. Note that the number of molecules aligned in this diagram is an exaggeration to emphasize alignment. In practice, only 0.1-1% of molecules are aligned (C).



The dependence of RDCs on $(3\cos^2\theta - 1)$ with respect to the external magnetic field means that they are averaged out in solution because of isotropic molecular tumbling (Figure 3.5B). Thus, RDCs are collected under anisotropic environments where proteins are placed in media that induces partial alignment. In these conditions, where roughly 0.1-1% of the protein molecules are aligned to the magnetic field, RDCs are not averaged out to zero (Figure 3.5C) (110, 111, 153). D_{NH} , RDCs describing the N-H amide bond vector, can by collected from IPAPtype HSQC experiments, which allow for reduced spectral overlap (111, 154). Two types of data sets are needed. The first is an in-phase (IP), which is the same as the HSQC discussed in Chapter 1 without proton decoupling to ¹⁵N during t_1 evolution. The second is an anti-phase (AP), where an extra INEPT step is used to transfer the anti-phase term to an in-phase term prior to ¹⁵N evolution. These two spectra are then used to derive two HSQC-type spectra (one by adding these two, the other by subtracting them). Both the IP and AP experiments are collected under two conditions, isotropic and in the presence of an alignment media. In the isotropic condition, the ¹⁵N chemical shifts are displaced by J_{NH} . With partial alignment, the chemical shifts are displaced by an additional, D_{NH} , the RDC (Figure 3.6).

Figure 3.6. Extracting couplings using IPAP-type experiments. Two experiments – an in-phase (IP) and anti-phase (AP) are collected to extract RDCs and to reduce spectral overlap. These are collected with proton coupling, resulting in splitting in the nitrogen dimension. Two HSQC-type spectra are derived from these as the sum (IP + AP) and difference (IP – AP). These are then overlaid, and the splitting corresponds to the coupling (A). In an alignment media, residual dipolar coupling can be determined by taking the difference between the splitting in alignment media and the scalar coupling determined from isotropic conditions in A (B).



Since RDCs report on the orientations of all internuclear vectors with respect to a single alignment tensor, RDCs are able to provide global orientational restraints, as discussed above. Furthermore, since RDCs reflect ensemble averaged dipolar couplings covering timescales up to milliseconds, they can also inform ensemble minimization. RDCs, for instance, were used to generate an ensemble of unliganded ubiquitin that covered the known conformational heterogeneity of ubiquitin bound to all of its known binding partners in the PDB (*123*, *139*).

Thus, in structure determination, RDCs provide orientational restraints that complement and supplement NOE distance restraints (<~5 Angstroms). In the study of protein dynamics, the power of RDC analysis lies on its ability to cover a wide range of timescales, including an inaccessible 'hidden' time, that are relevant to protein function, including molecular recognition and potentially enzyme kinetics.

Fingerprint spectra and detecting protein interactions. One of the simplest, and perhaps routine, NMR experiments done in most, if not all, biochemistry labs that work on proteins is an HSQC measurement, particularly a correlation between ¹H and ¹⁵N, typically referred to as a fingerprint spectrum (*155*). The details of the HSQC experiment have been discussed in Chapter 1 and a simplified schematic is again shown in **Figure 3.7**. In a ¹H, ¹⁵N-HSQC, all N-H correlations are observed, covering all backbone amide groups, as well as side chains from tryptophan, asparagine, glutamine, arginine, and lysine. The number of detectable peaks, therefore, can be predicted from the primary sequence of the protein. Prolines

contain a secondary imine group, where the side chain connects to the backbone nitrogen, and thus, do not show up in the HSQC. Providing backbone and sidechain information, this fingerprint spectrum is typically the first NMR spectra collected and can be used for assessing the viability of a system for NMR studies, the folding state of the protein, or for screening interactions. **Figure 3.7. Schematic for** ¹**H**, ¹⁵**N-HSQC.** The magnetization starts on the proton and is transferred, via scalar coupling to the attached ¹⁵N amide nitrogen. The chemical shift is evolved on the nitrogen before magnetization is transferred back to the proton for detection. The protein needs to, at least, be ¹⁵N-labelled.



Residue n-1

Residue n

A similar experiment, ¹H,¹⁵N-TROSY (for transverse relaxation optimized spectroscopy), has also been developed for larger systems (*34*). Larger molecules simply will have more signals causing spectral overlap and generally tumble slower in solution leading to a fast transverse relaxation rate translating to a broader linewidth and poorer signal-to-noise ratio. TROSY alleviates these issues by constructive use of the chemical shift anisotropy (CSA) and dipole-dipole (DD) coupling interference. In an HSQC that is collected without decoupling, the peaks will appear as a multiplet due to *J* coupling, with varying linewidths due to contributions from different relaxation mechanisms. TROSY selects for the slowest relaxing multiplet component, which is the sharpest peak, thus extending the applicability of NMR to larger systems at higher magnetic fields (**Figure 3.8**). The advantage of TROSY over HSQC is best observed with larger proteins at high magnetic fields.

Figure 3.8. HSQC versus TROSY peak. In practice, as discussed in Chapter 1, an HSQC is run with decoupling resulting in a single ¹H-¹⁵N cross peak (decoupled HSQC). Without decoupling, the peak is split into four corresponding to different relaxation rates due to combined effects of chemical shift anisotropy (CSA) and dipole-dipole coupling. This is illustrated in the coupled HSQC where the single peak in the decoupled spectrum is split into four with varying linewidths. The TROSY experiment selects for the peak in the lower right (TROSY).



The HSQC/TROSY peaks are assigned using a set of three-dimensional experiments for sequential assignment, such as an HNCA (correlates backbone nitrogen to its alpha carbon and that of the preceding residue) and HNCO (correlates backbone nitrogen to the carbonyl carbon of the preceding residue), among others. The assignment of the fingerprint spectrum is necessary for structure determination as mentioned in previous sections, but also provides further information that can be used to evaluate protein interactions. For instance, chemical shift perturbation in TROSY/HSQC can be used to screen for binders such as in protein-observed drug screens used in Chapter 5. Chemical shifts are sensitive to their molecular environment. As such, ligand or drug binding can be detected in HSQC/TROSY and manifests as peak shifts. Typically, residues exhibiting significant chemical shift perturbation are within or close to the binding pocket. Any allosteric effects of binding are also detected. Chemical shift perturbations are calculated as the Euclidean distance between the peaks, typically with a scaling factor for the ¹⁵N chemical shift. The scaling factor, α , is the ranges from 0.10 to 0.45, but is most commonly set to 0.14. This is based on the difference in the range of backbone nitrogen and amide proton chemical shifts (156).

$$CSP = \sqrt{\frac{1}{2} \left(\delta_H^2 + (\alpha \delta_N^2) \right)}$$

In addition to detecting an interaction, identifying binding sites, and identifying potential allosteric involvement, the fingerprint spectrum is also used to characterize binding affinities by NMR titration. It is worth noting that detection of interaction does not, technically, require any assignments as any difference in the unbound and bound spectra will indicate the presence of an interaction. However, to extract more useful information such as binding site information, an assignment is necessary. Performing NMR titration on an assigned spectrum also provides information that will support assignment of the bound form.

As with any titration method, small amounts of the ligand are titrated into the NMR tube with the protein sample and a HSQC or TROSY is collected for each titration point. The data can then be fit into an equation that describes the relationship between chemical shift and ligand concentration to derive the binding affinity constant, K_D (156). Assuming a two-state binding where the ligand/protein is either bound or free, the data can be fit to the following equation where P_T and L_T refer to the total concentration of the protein and the ligand, respectively.

$$\Delta \delta_{obs} = \Delta \delta_{max} \frac{\left\{ ([P]_T + [L]_T + K_D) - [([P]_T + [L]_T + K_D)^2 - 4[P]_T [L]_T]^{\frac{1}{2}} \right\}}{2[P]_T}$$

Here, $\Delta \delta_{max}$ refers to the maximum chemical shift difference upon saturation of the ligand, and $\Delta \delta_{obs}$ is the change in the observed chemical shift in the free state. During the fitting process, $\Delta \delta_{max}$ is typically fit as a local variable (per residue) as the true value may not be reached in the experiment, while K_D is fit globally.

Moreover, on top of binding affinity, the behavior of the peaks during the titration also reveals the timescale of ligand binding (**Figure 3.9**) (*156*, *157*). The equation above is used for titrations in fast exchange, but several tools have been developed to characterize binding affinities under any exchange regime (*158*). Finally, in the fast exchange, peaks move linearly. A non-linear plot can indicate multiple interactions, multiple binding sites, or other complicated mechanisms (*156*). Titrations may therefore provide a lot of information about the binding between the protein of interest and its binding partner.

Clearly, even the simplest and most routine of NMR experiments can provide significant information in the study of proteins and their interactions. Thus, the HSQC/TROSY experiments should be routine in any biochemistry lab that studies protein interactions.
Figure 3.8. Exchange regimes in NMR titration. Ligand binding can occur in the fast, intermediate, or slow exchange regime in NMR. A sample titration in both 2D (left) and the 1D projection (right) with increasing exchange from slow (top) to fast (bottom). In the fast regime, the peaks move smoothly and linearly in response to increasing amount of ligand. Meanwhile, in slow exchange, the peak corresponding to the free protein decreases as ligand is titrated. The peaks can be used to estimate the amount of protein bound. In the intermediate cases, the lineshapes are more complicated but are typically broadened. This image was adopted from Figure 1b,c of *Waudby, et al. (2016) Sci. Rep. 6:24826,* licensed under CC-BY-4.0. Image was cropped and panels rearranged and resized without further modification.



The drug discovery Swiss army knife. Borrowing the phrase from a recent publication, NMR can indeed be considered a 'Swiss army knife' in drug discovery with all the tools and experiments available from screening, to characterization of candidate drugs and their binding pose within the protein (*159*). One common application and one that will be applied in this work is protein-observed NMR using the simplest experiment described in the previous section – using the fingerprint spectrum to identify protein binders and to characterize the binding affinities, when possible (*156*). However, as mentioned, there are plenty of other applications that will briefly be mentioned here. Interested readers are encouraged, as always, to look at recent reviews that cover a vast majority of the methods described here, as well as others (*160–162*).

Fragment-based drug discovery is an approach that begins with identifying smaller molecule fragments, typically no bigger than 300 Daltons (*163–165*). Due to their smaller size, binding to target proteins is weak in the low micromolar to millimolar range. NMR is well suited for detecting even these weak interactions and is thus used in the identification of promising fragments. From these hits, additional fragments that can potentially be linked to the original fragments may be identified, or the fragment can be used as a backbone for drug development (**Figure 3.10**).

Figure 3.10. Principle of fragment-based drug discovery. Small fragments that weakly bind to a target protein can still be detected by NMR. In protein-observed fragment screening, three small molecules that perturb different parts of the spectra, and thus, bind to different parts of the protein, may be identified. These three molecules – the green triangle, purple circle, orange rectangle – individually will bind weakly to the protein. If their binding sites are close to each other, the fragments can be linked yielding a higher affinity molecule. NMR can be used in each step, from fragment identification, to validation and characterization of the high affinity molecule. The blue sphere represents a part of the protein and the white portions the druggable pocket.



In Structure Activity Relationship (SAR)-by-NMR, for instance, one approach is to use the fingerprint spectrum (*166*, *167*). Fragments that bind to the target protein will induce CSPs in their binding sites. Two fragments that induce CSPs in distinct sites may then be linked to yield a possibly higher affinity ligand, or a secondary screen may be performed on the complex of the protein and the first fragment. NMR is especially suited for this as it can detect these weak interactions and also provide binding site information simultaneously. Beyond fragment screening, NMR can also be used in high throughput screening (a smaller version is described in Chapter 5) especially when coupled with automation (*168*, *169*).

At the same time, NMR is a tool to characterizing hits in drug discovery. The ability of NMR to determine binding affinity has already been discussed in detail in the previous section, using an NMR-monitored ligand titration. On top of that, there are ligand-based or ligand-observed methods that allow accurate characterization of protein binding. Some of these include saturation transfer difference (NMR-STD) and WaterLOGSY (**Figure 3.11**). NMR-STD yields a difference spectrum of the ligand obtained from the difference between the reference spectrum of the ligand and a spectrum collected under saturation of receptor signals (*170, 171*). If the ligand is bound to the receptor, the perturbation on the receptor can be transferred to the ligand resulting in modulated signals. Signals of molecules that do not bind, or chemical entities within a ligand that do not interact with the receptor, will cancel out (**Figure 3.11A**). This method therefore informs chemical screening as well as chemical optimization, such as through structure-activity relationships (SAR).

water molecules to the ligand, which may occur directly or via initial transfer to the proteins and onto the protein-bound ligands (*172*, *173*). Molecules or fragments that interact with the protein appear as having negative signals (**Figure 3.11B**). Thus, similar to NMR-STD, WaterLOGSY can also be used for screening and optimization of small molecules.

Figure 3.11. Examples of WaterLOGSY and STD-NMR. STD-NMR was used to identify drug fragments that bind to the adenosine A2A receptor (A2AR). The difference spectrum, labelled STD, indicate binding if there is no loss of signal. Fragments 12, 13, 14, and 15 show strong binding, while fragment 11 is a weak binder at best. This image was adopted from Figure 4c of *Igonet, et al. (2018) Sci. Rep. 8:8142,* licensed under CC-BY-4.0. Image was cropped without further modification. (A). WaterLOGSY shows that compound 3344 binds to KRAS, as seen in the negative signals. Addition of the anti-RAS single chain variable fragment (scFv) blocks this binding. This image was adopted from Supplementary Figure 3E of *Bery, et al. (2018) eLife 7:e37122,* licensed under CC-BY-4.0. Image was cropped without further modification (B).



Thus, NMR finds an application in every aspect of drug discovery from molecule/fragment screening to optimization and characterization. Not to mention, atomic-level characterization of small molecules is also done by NMR.

In this chapter, some applications of NMR, particularly relating to the current investigations, are briefly discussed including some methods. NMR has been described as the 'Swiss army knife' in drug discovery, but its applications, clearly, go beyond drug discovery, and it should perhaps more aptly be the Swiss army knife of structural biology. The structural biology toolkit consists of ever improving methodologies that are able to characterize protein structure and dynamics, but NMR whether alone or in an integrative approach, remains uniquely flexible for various applications. As other techniques improve, so too shall NMR. Modern methods should be effective in studying larger proteins and protein complexes – both structures and dynamics. Computational methods, including Al-based algorithms, meanwhile, continue to improve virtual screening, but it is hard to imagine a drug discovery program that does not, at some point, employ NMR whether in the characterization of protein binding or the characterization of the small molecule.

CHAPTER 04

GENERAL PROTEIN BIOCHEMISTRY METHODS

AND RELATED ASSAYS

"At this point, it should be clear that every lab aiming at obtaining a recombinant protein in E. coli should have at least a stock of the BL21(DE3) strain and vectors of the pET series.

> Rosano and co. Protein Science 2019 (174)

This present work involves the study of three completely different protein systems. Paramount to any work relating to proteins – including drug discovery and structural research – is the expression, purification, and characterization of the purified protein product. This chapter summarizes some routine protocols used in the expression, purification, and characterization of PRL3, MDMX, and GMPK. A discussion of each of these systems is not covered in this chapter but will be described in detail in the subsequent chapters.

Expression and Purification of Recombinant Proteins. In the study of protein structure, regardless of the methodology, the common requirement is access to

pure, high quality protein products. The same is true for drug discovery, particularly in the *in vitro* screening and binding characterization. While there are several possible hosts including human cell cultures, insect cells, and yeast, the most common expression system is *E. coli*. Recombinant expression in *E. coli* has its advantages over the other expression systems (*175–177*). Bacteria have faster growth kinetics and high cell densities are easily achievable using controllable media. They are also more easily handled, and several well-characterized strains and plasmids have been developed for specific applications. Different rich media are available and relatively inexpensive, while chemically defined formulations are also accessible. Chemically defined media are necessary for NMR structure determination and other applications that require incorporation of stable isotopes (*178*). Common reagents used for labelling are listed in **Table 5**. **Table 5.** Common reagents used for stable isotope labeling in NMR.

Nuclei	Reagent			
² H	D ₂ O (replacing water in media)			
¹³ C	¹³ C-glucose*, ¹³ C-glycerol*			
¹⁵ N	¹⁵ NH₄CI			
¹⁹ F	Fluorinated amino acid analogs**, TFEA (Glu)***, TET (Cys)***			
¹³ C (ILV)	Labeled ketobutyrate and ketoisovalerate*			

* Some experiments may require selective carbon labeling.

** Using chemically defined media, during expression.

*** Chemical modification post-purification

In this work, all proteins were produced in a *BL21(DE3)* variant, *Rosetta(DE3)*. It is worth noting, however, that this system was chosen because of availability and the relatively good yield after the first pass, without need for significant optimization. For a protein that has not been purified before, any of the commonly available expression strains should work, but there is a selection for specific applications such as for systems with rare codons, those that require altered redox states for folding of proteins with disulfide bonds, or auxotrophs for incorporating non-canonical amino acids, among others (*174*). The *DE3* designation in these *E. coli* strains indicate that they carry the gene for the T7 RNA polymerase under the control of the lacUV5 promoter, a mutated version of the *lac* promoter in *E. coli* that works independent of any activator proteins and other *cis* elements (*179*). A few of the most common strains used for recombinant protein expression are summarized in **Table 6**.

In this study, the genes coding for the proteins of interest are all cloned in a modified pET28a plasmid that appends an N-terminal 6X-His tag and an TEV protease cleavage site (**Figure 4.1**). The pET family of vectors are under the control of the T7 promoter and the expression is induced by an allolactose analog, isopropyl ß-D-1-thiogalactopyranoside (IPTG). There are several available plasmids, and the construct may need to be optimized for soluble expression (*174*).

Strain	General Use			
BI 21(DE3)	General protein expression; good first strain for			
	optimization			
	General protein expression; Contains plasmid which			
BL21(DE3)	expresses high (pLysE) or low (pLysE) levels of T4			
pLysS/pLysE	lysozyme which inhibits the T7 RNA polymerase for			
	inhibiting basal expression			
Pacatta(DE2)	Contains pRARE, a plasmid that codes for tRNA for			
	low-usage codons in <i>E. coli</i> ; also has pLysS version			
	Has at least one mutation that prevents cell death			
C41(DE3)	associated with expression of toxic recombinant			
	proteins			
	Expression of toxic and insoluble proteins; contains			
Tuner(DE3)	mutated lac permease for linear control of expression			
	levels			
Origami(DE3)	Contains highly active active thioredoxin reductase			
	and glutathione reductase for proper folding			

Table 6. Some *E. coli* strains used for expression of recombinant proteins*.

* Table adopted from Plasmids 101 by Julian Taylor-Parker from the Addgene Blog (https://blog.addgene.org/plasmids-101-e-coli-strains-for-proteinexpression). **Figure 4.1. pET28a vector map**. A modified pET28a vector was used in all experiments in the current work. Some general features of pET28a vectors are shown, including the kanamycin resistance gene, *Lacl* gene under the *Lacl* promoter, the origin of replication (ori), and the cloning site where the gene of interest is placed under the regulation of the T7 promoter (A). The cloning site has been modified to include a cleavable N-terminal 6X-histidine tag. The tag can be removed using TEV protease (B).



In this work, PRL3 (4-159), MDMX (7-111), and GMPK (full-length) can all be easily expressed and purified under the conditions described without further optimization. The average yield for all three proteins is 30-50 mg/L of Luria-Bertani (LB) or M9 (after four-fold concentration from LB) medium. All genes coding for the proteins were codon-optimized and synthesized (Genscript, NJ) before being subcloned into the modified pET28a vector (**Figure 4.1**). All plasmids used in this work have been verified by Sanger sequencing (Eurofins, KY) using the T7/T7 reverse primers flanking the cloning site.

For expression, the plasmid was transformed into *Rosetta(DE3)* (EMD Biosciences, CA). Then, a standard/routine expression protocol was followed (**Figure 4.2**). Transformant colonies were used to inoculate an overnight culture in LB at 25°C for 16-20 hours with shaking. The following day, main cultures were started by inoculating fresh LB with enough overnight culture to yield a starting OD600 of no more than 0.10. This was incubated at 37°C with shaking until OD600 reached 0.60-0.80, upon which 1 mM of IPTG was added to induce protein expression. After 4-5 hours, cells were harvested at 6,000 x g and frozen at -20°C.

For the expression of ¹⁵N and/or ¹³C-labelled proteins, cells were pelleted upon reaching OD 0.60-0.80 and were washed with M9 salts. Cells were then resuspended in supplemented M9 medium and allowed to grow for an additional hour before inducing protein expression for 4-5 hours (**Figure 4.2A**).

The standard minimal M9 salt solution was used consisting of 33.7 mM Na₂HPO₄, 22.0 mM KH₂PO₄, 8.55 mM NaCl, at pH 7.2. The final medium was

143

supplemented with 1 mM MgSO₄, 1 μ g/L thiamine, and 10 ml/L of 100X DMEM Vitamin solution per L. The nitrogen and carbon sources used are 1 g/L NH₄Cl and 3 g/L glucose and are substituted with the labelled counterparts as necessary.

For the expression of ²H-labelled versions of the proteins, an adaptation protocol was used (**Figure 4.2B**). A modified M9, referred to as M9++, was also used. This medium is composed of 10.9 mM K₂HPO₄, 3.67 mM KH₂PO₄, 6.3 mM Na₂HPO₄, 1.4 mM K₂SO₄, 0.1% LB in D₂O, 1 mM MgSO₄. A trace element solution may also be added. For every 100 ml of M9++, 0.5 g NH₄Cl and 1.8 g of glucose is used and swapped with stable isotope labelled material as needed.

For a 100 ml final culture, 1 ml of LB is first inoculated with transformant colonies. After 5 hours, 100 μ l of this culture is used to inoculate 2 ml LB/D₂O, all of which is then used to inoculate a 20 ml M9++/ D_2O medium after an additional 5 hours. The M9++/ D_2O culture is allowed to grow for an additional 15 hours before again being transferred into the final 100 ml M9++/D₂O. Cells are induced with IPTG for expression after the initial 10 hours in the final culture, and protein expression is allowed to proceed for another 20 hours. This protocol was adapted from an ILV labeling scheme (180). For selective labeling of isoleucine/leucine/valine methyl groups, the precursors shall be added after the 10 hours in the final culture and incubated for an additional hour prior to protein expression.

Figure 4.2. Protein expression protocols. Standard protocols for the recombinant expression of proteins in *E. coli* are shown. For unlabeled proteins, rich media like LB and TB can be used. For ¹⁵N and/or ¹³C-labeled proteins, a '4X' protocol is typically used to obtain high density culture prior to induction of expression (A). An adaptation protocol is used for expression of ²H-labeled proteins or when precursors for selective labeling is necessary. This figure was taken, in its entirety and without modification, from *Cai, et al. (2021) J Biomol NMR* 75: 83-87 with publisher permission (license number: 5327751075366).



В	Cell adaptation			Cell growth	add precursor	s Induction			
$\int 100 \mu\text{L}$									
Vessel volume	15 ml	15 ml	250 ml	2.8 L	2.8 L	2.8 L			
Media	LB/H ₂ O	LB/D ₂ O	M9++/D ₂ O	M9++/D ₂ O	M9++/D ₂ O	M9++/D ₂ O			
Culture Volume	1 ml	2.0 ml	20 ml	100 ml	100 ml	100 ml			
Growth time	~5 h	~5 h	~15 h	~10 h	~1 h	~20 h			
Cell Density (OD ₆₀₀)	1-2	1-3	6-10	~8	~10	~20			
Temperature	37°C	37°C	37°C	37°C	37°C	25°C			

Protein purification in all cases was done using a combination of immobilized metal affinity chromatography (IMAC)/reverse IMAC and size exclusion chromatography (**Figure 4.3**). Native conditions were used except for MDMX which is expressed in inclusion bodies and requires purification in urea.

Cell pellets were thawed and re-suspended in 10-20 ml of lysis buffer (50 mM Tris, pH 8, 500 mM NaCl, 2 mM TCEP, 3 mM NaN₃) per gram of wet pellet. The lysis buffer was supplemented with PMSF at a final concentration of 1 mM. Cells were then lysed by sonification. The cells were kept on an ice-water bath to prevent unwanted heating during the sonification process. Cell debris was removed by centrifugation at 25,000 x g for 30 minutes. For MDMX, the cell debris was resuspended in 8 M urea, sonicated following the same protocol and again spun.

The supernatant was then loaded onto a column packed with Ni-NTA that has been extensively washed with lysis buffer, or with 8 M urea in the case of MDMX. To separate relevant fractions, the bound proteins were eluted using an imidazole gradient. The gradient length was set to 300 mins, with a flow rate of 2 ml/min, and a target concentration of 500 mM imidazole.

147

Figure 4.3. Basic chromatography purification protocol. His-tagged proteins can be purified by immobilized metal affinity chromatography, such as using Ni-NTA. If the tag is cleavable such as illustrated, then the tag is cleaved and a reverse step is done where flow through is collected. The protein can then be purified further by size exclusion chromatography. If the protein is in the insoluble fraction, a few extra steps for urea solubilization are necessary (yellow box).



The chromatogram for the first purification step of His-tagged PRL3 is shown in **Figure 4.4A**. Relevant fractions are identified by SDS-PAGE, pooled, mixed with sufficient TEV protease, and dialyzed against a TEV protease cleavage buffer (20 mM Tris, pH 7.2, 100 mM NaCl, 2 mM TCEP) overnight. This step combines removal of imidazole as well as cleavage of the N-terminal tag, and for MDMX removal of urea as well as refolding. MDMX samples are also diluted 1:2 in the cleavage buffer to reduce precipitation during the refolding process.

The next day, the mixture is again loaded onto a pre-equilibrated Ni-NTA column, and the flow-through is collected and concentrated. The concentrated protein is further purified by size-exclusion chromatography (SEC) (**Figure 4.4B**). The final buffer depends on the system and downstream applications. PRL3 was in Tris (30 mM Tris, pH 7.2, 100 mM NaCl, 5 mM TCEP, 3 mM NaN₃), MDMX in phosphate (30 mM Na₂HPO₄, pH 7.2, 200 mM NaCl, 2 mM TCEP, 3 mM NaN₃), and GMPK in MOPS buffer (30 mM MOPS, pH 7.2, 100 mM NaCl, 3 mM NaN₃).

Fractions containing the proteins of interest are identified and pooled and the protein concentration is adjusted to 100 μ M, aliquoted, and snap-frozen in liquid nitrogen for storage at -80°C. The concentrations are determined by measuring absorbance at A280 in 90% guanidine HCl, which is then converted to an estimated concentration using a theoretical extinction coefficient (Scripps Protein Calculator). Values used in this work are summarized in **Table 7**.

The quality of all proteins from each batch of purifications are assessed by SEC (**Figure 4.4B**), SDS-PAGE, and fingerprint NMR (discussed below). In all cases, it is evident that the final protein products are at least 95% pure.

Figure 4.4. Ni-affinity and SEC profiles for PRL3. Chromatogram for the first Ni-NTA purification step of PRL3. His-tagged proteins typically elute at ~20% (of 500 mM imidazole buffer) boxed in pink. The first peak corresponds to non-specific binding (A). After cleavage and a second Ni-NTA step, PRL3 is further purified by size exclusion chromatography (B).



Drotoin	Residues in	MW	nl	Ext. Coeff
Protein	Construct	(kDa)	р	(M⁻¹ cm⁻¹)
PRL3	4-159	17.5	9.06	19,060
GMPK	1-197 (full)	21.7	6.55	6,400
MDMX	7-111	11.7	7.89	6,400

 Table 7. Relevant properties of PRL3, GMPK, and MDMX for purification

Assignment of backbone resonances. The fingerprint spectra (HSQC or TROSY) were collected with proteins that were typically concentrated to 0.5 - 1mM. As mentioned in the previous chapter, there are a routine set of 3D experiments used for backbone resonance assignment. In the case of MDMX and PRL3, these were mostly accomplished using 3D-HNCA (Figure 4.5). In this experiment, magnetization is transferred from the proton to the ¹⁵N and then to the ¹³Cα and back to ¹⁵N and to the proton for detection. The chemical shift is evolved in all these nuclei to yield a 3D spectrum. The ¹⁵N is coupled to the C α of the current and preceding residue. Typically, the C α of the current residue is closer to the ¹⁵N and the two peaks can therefore be distinguished based on their intensity. Some residues have unique C α chemical shifts and can be used to guide resonance assignments. For instance, glycines have really low C α chemical shifts at about 45 ppm compared to the Ca chemical shift range of 52-63 ppm for the other residues. This information is used to sequentially assign the backbone. An illustration of the process is shown for GMPK (Figure 4.6)

Additional experiments such as CC(CO)NH-TOCSY were also used to confirm assignment (*181*, *182*). In this experiment, magnetization is transferred from side chain protons to their attached ¹³C. Isotropic mixing allows transfer of this magnetization among the carbon nuclei, which is then transferred to the carbonyl, and then to the amide group, and finally the amide proton for detection. The result is a 3D spectrum where an NH tower contains the carbon chemical shifts of the side chain of the preceding residue. Coupled with an HNCA, the C α chemical shifts as well as the spin system can be verified.

154

Figure 4.5. Schematic for 3D-HNCA. The magnetization starts on the proton and is transferred, via scalar coupling to the attached ¹⁵N amide nitrogen. Then, it is transferred to the ¹³C α via N-C α scalar coupling before being transferred back to ¹⁵N and then proton for detection. The chemical shift in ¹H, ¹⁵N, and ¹³C yielding a 3D dataset. As illustrated, amide nitrogen is coupled to the C α of the current residue (one bond) and that of the previous residue (two bonds). The coupling constants (¹J and ²J) are different and the intensities of the peaks in the HNCA strip are also different. Typically, the C α directly linked to the amide nitrogen (residue *n*) will have a higher intensity, permitting sequential assignment. The protein needs to be ¹⁵N- and ¹³C-labelled.



Residue n-1

Residue n

Figure 4.6. Part of the HNCA of GMPK. A peak in the HSQC spectrum (A) corresponds to a correlation between an amide proton and nitrogen. For each resonance in the backbone, there is a corresponding strip in the HNCA. The two peaks correspond to the two C α carbons that are coupled to the amide nitrogen. The more intense peak corresponds to residue *i*, while the less intense peak corresponds to residue *i*, while the less intense peak corresponds to residue *i*. In sequential assignment, the information on the C α carbons is used and linked together. A linked fragment for GMPK is shown as an example where the peaks can be followed from residue 33 to 40 (B). This experiment was done at 600 MHz and 25°C with 1024, 100, 60 points in the direct (*t*₃) and indirect dimensions (*t*₂, *t*₁), respectively with 64 scans.



In the present work, these experiments were used to assign as much of the backbone resonances for the three proteins under study. When available, existing backbone assignments supported the analysis, and the experimental data was used to confirm the assignments.

The assigned backbone resonance for MDMX, PRL3, and GMPK are shown in **Figures 4.7-4.9**. All protein samples were concentrated to 500 μ M to 1 mM in their respective buffers supplemented with 10% D₂O. The assigned backbone resonances shown are for apo PRL3, apo GMPK, and nutlin-bound MDMX. Experiments were carried out at 25°C at 600 MHz. The spectra were recorded with 1024 and 100 complex points in the direct (*t*₂) and indirect (*t*₁) dimensions, respectively, with 8 scans.

Figure 4.7. ¹H, ¹⁵N-TROSY for apo PRL3 with backbone assignments.


Figure 4.8. ¹H, ¹⁵N-HSQC for MDMX-nutlin3a complex with backbone assignments.



Figure 4.9. ¹H, ¹⁵N-HSQC for GMPK with backbone assignments.



Enzyme activity/inhibition assays. To ascertain whether recombinantly expressed and purified proteins are properly folded and functional, activity assays are necessary. PRL3 has phosphatase activity while GMPK is a kinase. These functions can be ascertained by enzymatic assays. Functional characterization of the GMPK construct used here has been verified (by Sabo lab) and not further discussed in this work (*183*). In this work, activity and inhibition assays were used to confirm the functional form of PRL3, and the same assays were used for all screening and validation experiments.

In the in vitro enzymatic assays, a synthetic substrate, 6,8-Difluoro-4-Methylumbelliferyl Phosphate or DiFMUP (Life Technologies, CA), was used. PRL3 does not show significant activity against the other common synthetic substrate, p-nitrophenyl phosphate (pNPP). For the activity assays, the reaction mixture consisted of 2.5 µM of recombinant PRL3 and 30 µM of DiFMUP. The reaction buffer contains high concentration of reducing agent, either 10 mM DTT or 10 mM TCEP, to ensure that the active site cysteine is fully reduced. PRL3 is first incubated with this reducing buffer prior to activity or inhibition assays. For inhibition assays, 1 mM of the test drug is added to PRL3 and incubated at room temperature for 15 minutes. The reaction is initiated by addition of DiFMUP and is allowed to proceed for 20 minutes prior to measurement of fluorescent intensities in a BioTek plate reader with Ex/Em of 360/460 nm. The exact same set-up is used for dose-dependence analysis with varying drug concentration. The IC_{50} was calculated from a hill-slope model shown below with Max and Min referring to the asymptotes at maximum or minimum inhibition and *n* is the Hill slope.

$$I(x) = Max - \frac{Max - Min}{1 + \left(\frac{x}{IC_{50}}\right)^n}$$

An identical set up was used for kinetics analysis (**Figure 4.10**), but instead of endpoint measurement, points were collected every 20 seconds within the 20-min time frame. There is a roughly 30-second delay between addition of the substrate and the first reading due to mechanical limitations of the plate reader. Initial velocities, v_o were determined, and Lineweaver-Burke analysis was done using in-house scripts to determine the maximum velocity, V_{max} , and the Michaelis constant, K_m .

$$\frac{1}{v_0} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$$

PRL3 in phosphate buffer with no reducing agent was used as the negative control, while DMSO, DMF, or buffer in place of the test compound were used as a positive control, as appropriate. All drugs from the FDA panel screen are dissolved in DMSO. Some compounds in validation assays were dissolved in DMF (thienopyridone) or in buffer (olsalazine) and so the corresponding proper controls are used.

All experiments are done in at least triplicates and statistical analyses were done using statistics packages from SciPy.

Figure 4.10. Sample PRL3 kinetics assay. DiFMUP (structure inset) is used as a PRL3 substrate for *in vitro* activity and inhibition assays. Shown is a representative trial of a kinetics assay to validate activity of recombinantly expressed PRL3. The activity is measured in terms of relative fluorescence units (RFU).



ITC: The Gold Standard for Binding Studies. Arguably, the most basic function of a protein is molecular recognition – whether it be interacting with binding partners in complexes or recognizing a substrate prior to the catalytic step. Thus, an accurate determination of binding is necessary. In this work, since MDMX does not have any enzymatic activity, its binding to nutlin-3a was used to assess its function. This was especially important since not only was MDMX recombinantly expressed in this study, but it was also purified in the unfolded state, prior to folding. In this study, isothermal titration calorimetry (ITC) was used to characterize this binding. Additionally, ITC was also used for PRL3 to characterize its binding to potential inhibitor, olsalazine, and also for GMPK and its interaction with one of its substrates, GMP.

To date, ITC is still considered the gold standard for any binding studies (*184*). ITC relies on the detection of the heat released or absorbed when two molecules interact. In the current study, this is the heat associated with the interaction between a protein and a ligand or a substrate. One of the advantages of ITC is that it is label-free, requiring absolutely no chemical modification of either the protein or the ligand being studied.

The heat associated with the binding event is measured by using two cells at the same temperature: the sample cell containing the protein under investigation, and a reference cell that contains a matched buffer. The calorimeter senses any temperature difference between the two cells – such as in response to heat released during the binding between a protein and a ligand – and compensates for the difference to return them to equal temperature.

In the current study, the proteins are all placed in the sample cell at concentrations ranging from $30 \ \mu\text{M}$ to $60 \ \mu\text{M}$ and the syringe is filled with the titrant at 10X concentration of sample cell. For instance, olsalazine was titrated to PRL3 (**Figure 4.11**), GMP to GMPK (**Figure 4.12**), and nutlin-3a to MDMX (**Figure 4.13**). Buffer matching was ensured by dialyzing all samples to the ITC buffer and using the same ITC buffer to dissolve or dilute the titrants. The ITC buffer was also used to fill the sample cell. As a control, the titrant is also titrated into the buffer.

In an exothermic binding, as was the case for all three systems, a downward peak arises from each injection. The initial heat changes are the largest as there are more available receptors for the titrant to bind to, and this decreases as the receptors are saturated. After the injection, the ligand binds to the protein, heat is released and measured by the instrument, and the system is allowed to return to equilibrium before the next injection. Eventually, an injection does not produce any more measurable change indicating saturation of the system. The peaks are then integrated, and the molar ratio is plotted against the enthalpy change, which can be fit to a model to determine binding affinities. In the case of the three systems under study, all binding events were expected to be one-to-one, so a single binding site model was used (*185–187*).

The binding constant is defined based on the fraction of sites occupied by the ligand X with [X] denoting free ligand concentration and X_t , M_t are the total concentrations of the ligand and the macromolecule, respectively.

$$K = \frac{\Theta}{(1 - \Theta)[X]}$$

$$X_t = [X] + \Theta M_t$$

Meanwhile, the total heat content Q contained in the active cell volume V_0 given that a fraction of the sites, Θ , are occupied can be expressed as follows where ΔH is the enthalpy of ligand binding:

$$Q = \Theta M_t \Delta H V_o$$

Combining these expressions result in an expanded expression for Q which can then be used to calculate the heat, Q(i), associated with each injection *i*:

$$Q = \frac{M_t \Delta H V_o}{2} \left(1 + \frac{X_t}{M_t} + \frac{1}{KM_t} - \sqrt{\left(1 + \frac{X_t}{M_t} + \frac{1}{KM_t}\right)^2 - \frac{4X_t}{M_t}} \right)$$

This analysis will provide the binding affinity, K or K_A , which is easily converted to a dissociation constant, K_D .

$$K_A = \frac{1}{K_D}$$

Figure 4.11. Olsalazine binding to PRL3 by ITC.



Figure 4.12. GMP binding to GMPK by ITC.



Figure 4.13. Nutlin-3a binding to MDMX by ITC.



While ITC remains the gold standard in the measurement of binding affinities, several other methods are available. These include surface plasmon resonance (SPR), microscale thermophoresis (MST), and biolayer interferometry (BLI) (*188–190*). Compared to ITC, these three techniques use smaller amounts of samples. However, ITC is label-free (as opposed to MST) and does not require sample immobilization (as opposed to SPR and BLI).

In designing ITC experiments, some considerations must be made, particularly in deciding the concentrations of the protein and the ligand to be used. Information on the expected K_D significantly helps in planning experiments. Interested readers are directed to manuscripts that tackle experimental design in detail (191–193).

In this chapter, the routine experiments used in the subsequent three chapters are discussed. The success of the three areas of research presented in Chapters 5-7 rely on the production of functional proteins and the methods presented here are not, in anyway, exhaustive, but are arguably routine in most biophysics or protein biochemistry laboratories.

CHAPTER 05

IDENTIFYING SMALL MOLECULE INHIBITORS AGAINST ONCOGENIC PHOSPHATASE, PRL3

"NMR has become a gold standard method in drug design due to its speed, simplicity, and reproducibility."

Emwas and Company *Molecules 2020 (162)*

The ability of NMR to provide residue-level information on protein interactions, discussed briefly in previous chapters, makes it a powerful tool to validate the interaction between proteins and other molecules, such as candidate compounds or fragments in drug discovery. For certain target protein systems, protein-observed NMR can be a quick method to screen for molecules that directly interact with a protein or a way to validate and support the results of *in vitro* drug screens. One such application is covered in this chapter. While the ultimate goal for this chapter is to identify inhibitors of PRL3, it is illustrated how NMR is used, and is essential, to support drug discovery programs.

Targeting protein phosphorylation. Protein phosphorylation is a posttranslational modification (PTM) involved in the regulation of cellular signaling modulating several biological processes (*194–200*). It is a reversible mechanism – in a way, acting as a switch – that alters properties of proteins including cellular localization, activity, stability, and interactions ultimately affecting signaling pathways in which these proteins are crucial participants (**Figure 5.1**) (*199*). An appropriate level of phosphorylation is crucial to sustaining normal cellular functions, while aberrant phosphorylation of proteins is implicated in many diseases including cancers, inflammatory diseases, and diabetes and obesity, among others (*197, 198, 201–204*). This PTM is regulated by two classes of enzymes, phosphatases and kinases, that work concert with each other. Thus, aberrant phosphorylation is ultimately the result of the dysregulation of kinases and/or phosphatases such that both are significant potential therapeutic targets for diseases characterized by abnormal phosphorylation levels. **Figure 5.1. Regulation of cellular activities by phosphorylation.** Phosphorylation affects cellular signaling pathways by altering the 'function' of proteins. This can occur via several mechanisms such as altering stability, activity, half-life, localization, and interactions. This figure was taken, in its entirety and without modification, from *Humphrey, et al. (2015) Trends Encocrinol Metab 26:* 676-687 with publisher permission (license number: 5327950835923).



Trends in Endocrinology & Metabolism

In cancers, protein phosphorylation has been shown to influence critical processes including cellular growth, proliferation, survival, and cell division (*198*, 205–207). There is no surprise, then, that kinases have been the target of several drug discovery programs – to a point that protein kinases may be considered the major drug targets of the 21st century (*208–211*). The success of Gleevec/Imatinib, the first molecule that is FDA-approved to specifically target a protein kinase, has paved the way for the development of more protein kinase-directed therapeutics (*208, 211–213*). To date, there are 89 drugs that have been approved by the FDA for clinical use for various indications, including anti-cancer drugs (according to a database compiled by the Medical Research Council Protein Phosphorylation and Ubiquitylation Unit at the University of Dundee). Clearly, targeting protein phosphorylation by modulating kinase activity is a viable strategy to developing therapeutics.

The curious case of phosphatases. Given the success in targeting kinases, one might assume that a similar level of success is attained in targeting phosphatases. After all, phosphorylation is a reversible PTM: kinases attach a phosphate group, while phosphatase remove the phosphate group (*195*, *199*). Yet, there has been a significant lag in phosphatase drug discovery and there currently are no FDA-approved inhibitors that directly target phosphatases (*202*, *214*, *215*). Calcineurin, perhaps, presents a special (and certainly, interesting) case as it is inhibited by the complex formed by FK506-binding protein (FKBP) and tacrolimus, now classified as a molecular glue (*216*).

Using the 'writer-reader-eraser' view of protein phosphorylation, kinases *write* the phosphorylations while phosphatases *erase* them. In other words, phosphatases are merely responsible for turning signals off and returning pathways to their basal levels; they act as 'housekeeping' genes (*217*, *218*). As recalled by Tonks, this misconception ultimately "promoted an emphasis on the study of kinases, in the signal transduction community, and a somewhat dismissive attitude to phosphatases" (*218*). In the same way, phosphatases have historically been viewed to be less specific and less tightly regulated than kinases, leading to a further gap in the study of and drug discovery against phosphatases and kinases (*219*, *220*).

In recent years, however, there is an increasing interest in identification of small molecules that directly inhibit phosphatases. Protein phosphorylation – and therefore, aberrant phosphorylation in the diseased states – is regulated by the concerted action of kinases and phosphatases (*218*, *219*, *221*). As such, both protein classes, indeed, deserve at least equal attention.

Phosphatase: from 'undruggables' to 'undrugged'. Beyond overcoming the misconceptions mentioned above, phosphatases have also eluded traditional active site inhibition. The active sites of phosphatases are strongly conserved with a high preference for negative charges as they have evolved to recognize phosphorylated substrates (*215*, *220*, *222*). This has earned phosphatases the stigma of being 'undruggable' (*214*, *218*, *219*, *223*). In contrast, kinase drug discovery has initially focused on targeting the ATP-binding pocket (*195*, *209*). In

addition to this strategy, however, other less conserved pockets have been identified within kinases towards the development of several classes of inhibitors (209, 224–227). Continued investment in phosphatase drug discovery, nonetheless, have allowed them to begin to shed the stigma (228–230). As further evidence of the direct and significant roles of phosphatases in diseases are accumulated, more research into their functions, structures, and interactions only continue to drive their transition from 'undruggable' to 'undrugged.'

Clinically relevant phosphatases. Several protein phosphatases, specifically protein tyrosine phosphatases, have been identified as potential therapeutic targets (*222*, *228*, *231*–*235*). While it is not the purpose of this section/chapter to discuss them in detail, a few of the well-characterized and highly sought-after phosphatase targets are briefly highlighted here.

PTP1B is perhaps the most well-studied and well-characterized phosphatase and indeed, the first protein tyrosine phosphatase (PTP) to be purified and structurally characterized (accounting for more than 50% of the PTP structures available in the PDB) (236–238). Some of the endogenous substrates of PTP1B include JAK2, insulin receptor (IR), and insulin receptor substrate 1 (IRS1) making it a pivotal regulator of the insulin and leptin signaling pathways (237, 239, 240). The roles of PTP1B in insulin sensitivity and attenuation of leptin actuation, and ultimately in the regulation of obesity and diabetes, were shown in mice driving further interest in drug development against PTP1B (241–243). Recently, PTP1B has also been implicated in cancers increasing interest in the identification PTP1B-

directed inhibitors (244–246). According to the PHAROS database, there are currently upwards of 700 active compounds against PTP1B, but none currently approved for any indication.

An argument can be made the Src homology 2 domain-containing protein tyrosine phosphatase (SHP2) is the second most studied PTP. Among the structures of the PTP catalytic domains in the PDB, SHP2 accounts for roughly 10%, second only to PTP1B. SHP2 is the first identified proto-oncogene in leukemia (*228, 247, 248*). Mutations that lead to the hyperactivation of SHP2 are associated with juvenile myelomonocytic leukemia, B-cell acute lymphoblastic leukemia, and acute myelogenous leukemia (*249*). The activation of SHP2 consequently leads to activation of the mitogen-activated protein kinase (MAPK) pathway, which plays a role in regulation of cellular proliferation and resistance to apoptosis (*250*). Perhaps as expected, SHP2, similar to PTP1B, has several (>90) active ligands according to PHAROS but also has no approved inhibitor for any indication.

Other protein tyrosine phosphatases that are currently of interest to oncology drug discovery include PTP4A3/PRL3 (further discussed in the rest of the chapter), dual specificity phosphatases (DUSP), DUSP1 and DUSP6, which are involved in the regulation of MAP kinases, and cell division cycle 25 (CDC25) which are known to regulate checkpoint response to DNA damage, among several others (*202*, *228*, *229*, *235*, *251*).

PRL3 as a significant target in cancers. PRL3 or PTP4A3 is a member of the phosphatases of regenerating liver family, along with PRL1 and PRL2, and are so named because they were discovered as being strongly up regulated in regenerating rat liver (*252*). In colorectal cancer patients, it was observed that PRL3 was significantly overexpressed in metastases relative to primary tumors sites (*253*). This brought prominence to PRL3 as a potential oncogene and therapeutic target in cancers. Since then, PRL3 expression has been noted in several other cancer types – including leukemia, gastric, liver, breast, and ovarian cancers among others – and has been typically associated with poor patient prognosis (**Figure 5.2**) (*254–264*). Clearly, there is undeniable evidence on the role of PRL3 in cancer progression, particularly in metastasis and related pathways (*265, 266*).

Figure 5.2. PRL3 expression correlates with poor patient prognosis. Survivability plots for colorectal (left) and late-stage breast (right) cancers classified based on the expression level of PRL3. High PRL3 expression in both cases correlate with low patient survival. Plots generated from data from the Human Protein Atlas (proteinatlas.org).



In vitro and animal model studies and network analyses have also shown that PRL3 regulates other cancer-associated pathways including cellular migration, invasion, and proliferation, survival and cell cycle progression, and angiogenesis among others (*265–274*). The exact mechanism by which PRL3 participates in these pathways are not yet fully elucidated but some knowledge on direct and indirect substrates are beginning to emerge (*266, 275*). For instance, there is evidence that PRL3 modulates the activation of Src, either directly or by modulating C-terminal Src kinase (CSK), which then drives migration and progression in T-cell acute lymphoblastic leukemia (*276*).

Overall, there is insurmountable evidence that PRL3 plays a significant role in the regulation of oncogenic pathways, particularly in disease progression and metastasis. This makes PRL3 a significant therapeutic target in several cancers.

PRL3 vs. PTPs: sequence and catalytic mechanism. PRLs belong to the superfamily of protein tyrosine phosphatases (PTP) and are characterized by two signature motifs, the P-loop or CX₅R and the WPD loop (**Figure 5.3**). PRL3 has 79% homology to PRL1 and 76% homology to PRL2 (*260*). The PRL family is also characterized by a polybasic region and a prenylation motif towards the C-terminus (*277–280*). Similar to other PTPs, cysteine (C104 in PRL3) and arginine (R110 in PRL3) in the CX₅R and aspartate (D72 in PRL3) in the WPD loop are involved in the catalytic activity of PRL3 (*277*). The catalysis step in PRLs is also very similar to other phosphatases (**Figure 5.4**). The active site cysteine (C104) acts as a nucleophile during the cleavage step resulting in the formation of a thiophosphoryl

intermediate. The arginine residue (R110) stabilizes substrate binding and thiophosphoryl intermediate by interacting with the phosphate moiety. The second aspartate residue in the WPD loop (D72) acts as a general acid by donating a proton to the substrate during the cleavage step and as a general base by activating a water molecule during the hydrolysis step. Thus, similar to other PTPs, the conserved loops are very critical to PRL activity.

The WPD loop of PRL3 is followed by three consecutive prolines (**Figure 5.3**). This feature is conserved among PRLs but not other PTPs. PRL2 has three and PRL1 has two prolines following the WPD loop. These prolines may affect the motion of the WPD loop as these additional prolines may induce conformational restraints (*280*).

Figure 5.3. PRL3 sequence and some notable features. The amino acid sequence of PRL3 and its how it compares to other phosphatases. The conserved motifs, the WPD loop and the P loop, are shaded blue, with the active site residues marked by an asterisk. The polybasic region is in orange while the prenylation motif is in green. The three prolines following the WPD loop are also highlighted (red box).

1	MARMNRPAPVEVSYKHMRFLITHNPTNATLSTFIEDLKKYGATTVVRVCEVTYDKTPLEKDG	62
177	NGDFNWIVPGKFLAFSGPHPKSKIENGYPLHAPEAYFPYFKKHNVTAVVRLNKKIYEAKRFTDAG	237
18	${\tt DGSGCYSLPSQPCNEVTPRIYVGNASVAQDIPKLQKLGITHVLNAAEGRSFMHVNTNANFYKDSG}$	82
110	LNRVMEKGSLKCAQYWPQKEEKEMIFEDTNLKLTLISEDIKSYYTVRQLELENLTTQETR-EILH	173
22	DLDLTYIYPN-IIAMGFPAERLEGVYRNNIDDVVRFLDSKHKN-HYKIYNLCAERHYDTAKFN	82
	* * *	
63	ITVVDWPFDDGAPPPGKVVEDWLSLVKAKFCEAPGSCVAVHCVAGLGRAPVLVALALIESG-M	124
238	FEHYDLFFIDGSTPSDNIVRRFLNICENTEGAIAVHCKAGLGRTGTLIACYVMKHYRF	299
83	ITYLGIKANDTQEFNLSAYFERAADFIDQALAQKNGRVLVHCREGYSRSPTLVIAYLMMRQKM	145
174	FHYTTWPDFGVPESPASFLNFLFKVRESGSLSPEHGPVVVHCSAGIGRSGTFCLADTCLLLMDKRKDPSSV	244
83	CRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFL	146
	WPD Loop P Loop	
125	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> <mark>YRPKQRLR</mark>	159
125 300	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> <mark>YRPKQRLR</mark> THAEIIAWIRICRPGSIIGPQQHFLEEKQASLWVQGDIFRSKLKNRPSSEGSINKILSGLD	159 360
125 300 146	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> <mark>YRPKQRLR</mark> THAEIIAWIRICRPGSIIGPQQHFLEEKQASLWVQGDIFRSKLKNRPSSEGSINKILSGLD DVKSALSIVRQNREIGPNDGFLAQLCQLNDR	159 360 176
125 300 146 245	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> <mark>YRPKQRLR</mark> THAEIIAWIRICRPGSIIGPQQHFLEEKQASLWVQGDIFRSKLKNRPSSEGSINKILSGLD DVKSALSIVRQNREIGPNDGFLAQLCQLNDRDLCQLNDRDLKKVLLEMRKFRMGLIQTADQLRFSYLAVIEGAKFIMGDSSVQDQWKELS	159 360 176 295
125 300 146 245 147	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> YRPKQRLR	159 360 176 295 208
125 300 146 245 147	KYEDAIQFIRQKRRGAINSKQLTYLE <mark>K</mark> YRPKQRLR	159 360 176 295 208
125 300 146 245 147	KYEDAIQFIRQKRRGAINSKQLTYLE YRPKQRLR THAEIIAWIRICRPGSIIGPQQHFLEEKQASLWVQGDIFRSKLKNRPSSEGSINKILSGLD DVKSALSIVRQNREIGPNDG	159 360 176 295 208
125 300 146 245 147	KYEDAIQFIRQKRRGAINSKQLTYLEKYRPKQRLR	159 360 176 295 208
125 300 146 245 147 160 361	KYEDAIQFIRQKRRGAINSKQLTYLEKYRPKQRLR	159 360 176 295 208
125 300 146 245 147 160 361 177	KYEDAIQFIRQKRRGAINSKQLTYLEKYRPKQRLR	159 360 176 295 208
125 300 146 245 147 160 361 177 296	KYEDAIQFIRQKRRGAINSKQLTYLEKYRPKQRLR	159 360 176 295 208
	177 18 110 22 63 238 83 174 83	177 NGDFNWIVPGKFLAFSGPHPKSKIENGYPLHAPEAYFPYFKKHNVTAVVRLNKKIYEAKRFTDAG 18 DGSGCYSLPSQPCNEVTPRIYVGNASVAQDIPKLQKLGITHVLNAAEGRSFMHVNTNANFYKDSG 10 LNRVMEKGSLKCAQYWPQKEEKEMIFEDTNLKLTLISEDIKSYYTVRQLELENLTTQETR-EILH 22 DLDLTYIYPN-IIAMGFPAERLEGVYRNNIDDVVRFLDSKHKN-HYKIYNLCAERHYDTAKFN 63 ITVVDWPFDDGAPPPGKVVEDWLSLVKAKFCEAPGSCVAVHCVAGLGRAPVLVALALIESG-M 238 FEHYDLFFIDGSTPSDNIVRRFLNICENTEGAIAVHCKAGLGRTGTLIACYVMKHYRF 83 ITYLGIKANDTQEFNLSAYFERAADFIDQALAQKNGRVLVHCRAGYSRSPTLVIAYLMMRQKM 174 FHYTTWPDFGVPESPASFLNFLFKVRESGSLSPEHGPVVVHCSAGIGRSGTFCLADTCLLLMDKRKDPSSV 83 CRVAQYPFEDHNPPQLELIKPFCEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFL WPD Loop P Loop

Figure 5.4. The catalytic mechanism of PRL3. The active site loops, WDP and P loop, are shown along with the three required active site residues. The dephosphorylation of a generic substrate is shown.



Meanwhile, in contrast to other PTPs, the PRL family members contain an alanine residue following the arginine residue in the CX₅R loop. In most PTPs, this position is occupied by a serine or a threonine residue, which may play an important role in the hydrolysis of the phosphoenzyme intermediate (*281*). Consistent to this, an increase in phosphatase activity is shown in PRLs upon replacing this alanine with a serine (A111S in PRL3) (*275, 281*). Furthermore, the CX₅R loop of the PRL3 family is highly hydrophobic due to residues between the catalytic residues (VAGLG) compared to other PTPs (**Figure 5.3**). The differences near this loop may confer a unique specificity to PRLs which could also be leveraged to developing selective PRL inhibitors.

Unique to PRLs are the polybasic region and prenylation motif (**Figure 5.3**). The prenylation motif at the C-terminus, also known as the CAAX motif, is a mark for prenylation and PRLs are typically farnesylated in this region allowing them to localize in the plasma membrane (*282*). The polybasic region that immediately precedes this prenylation motif, meanwhile, facilitates the binding of PRLs to the membrane by interacting with the negatively charged phospholipids in the membrane (*256, 281*).

The structure of PRL3 possess a similar general conserved architecture as other PTPs – a central β -sheet flanked by helices (**Figure 5.5**) (*283*). That said, while the general architecture is conserved loop orientations in PRL3 are significantly different from other phosphatases. Number, orientation, and lengths of secondary structures are also different in PRL3 (*278*). These differences might be leveraged for development of specific drugs. In the meantime, additional differences in
dynamics that may be related to these structural features will be further described towards the end of this chapter.

Figure 5.5. Structure of PRL3. The structure of PRL3 possesses similar general features as other PTPs with central sheets flanked by helices. PRL3 also exhibits a transition from the open (left) to closed (right) transition that is characterized by the movement of the active site loops (in red and orange). (PDB: 1V3A, 2MBC)



Current inhibitors of PRL3. The role of PRL3 in various cancers indicate that there is a significant need for the development of specific inhibitors. Currently, several inhibitors have been identified through high-throughput screening, virtual screening, and structure activity relationship (SAR) studies (**Figure 5.6**). Through a high-throughput screen of the Roche chemical library, thienopyridone was identified (*284*). Thienopyridone was demonstrated to significantly inhibit tumor cell anchorage-independent growth in soft agar and migration (*284*). A photooxygenation product of thienopyridone, iminothienopyridinedione (ITP), was developed through a SAR study and was shown to have anti-tumor activity against drug-resistant ovarian cancer in a murine xenograft model (*264, 285*). Both inhibitors show specificity towards PRLs, and specifically PRL3. However, thienopyridone is associated with toxicity and ITP is not very stable chemically, necessitating the need to find much better inhibitors that can be used in the clinic (*264, 284*).

Figure 5.6. Some available small molecule inhibitors of PRL3.



Rhodanine and derivatives were also identified as potential inhibitors through high-throughput screening. After SAR studies, CH707 and BR-1 were identified and are able to recover phosphorylation levels of potential PRL3 substrates and also suppress migration and invasion in cancer cells (*286*). Analog 3 has been developed based on virtual screening and SAR studies and shows inhibition in the migration of PRL3-expressing cells in a dose-dependent manner (*287*). Pentamidine, an FDA-approved drug, is also proposed as an inhibitor of PRL3 and shows inhibitory activity against the growth of multiple cancer cell lines and human melanoma WM9 cells in nude mice (*288*). A few natural products have also been identified as potential PRL3 inhibitors although their specificity remains to be examined (*289–291*). A virtual screening was also able to identify a PRL1 trimer formation inhibitor, compound 43, although evidence for trimerization of PRL3 is currently not available (*292, 293*).

Overall, while there are several proposed inhibitors for PRL3, issues with chemical stability, specificity, and toxicity limit their applicability in the clinical setting. Furthermore, these inhibitors do not have validated binding mechanisms to PRL3. Thus, to develop PRL3-directed inhibitors, there needs to be further studies in characterizing these compounds or additional inhibitors need to be identified.

PRL3 is a monomer. One of the members of the PRL family, PRL1 has been shown to form trimers under certain conditions (294). Thus, it is of interest to verify the oligomeric state of PRL3. Under the conditions used in all experiments in this

body of work, it is clear that PRL3 is a monomer. First, size exclusion chromatography (**Figure 4.4B**) indicates that the current construct of PRL3 used covering residues 4-159 is roughly 17 kilodaltons based on elution volume. The column used in purifying PRL3 was calibrated using a gel filtration standard (Bio-Rad, CA). To further confirm this, the TRACT NMR experiment was used. TRACT permits the estimation of a protein's rotational correlation time, τ_c (295). Based on TRACT, PRL3 has a tumbling time of about 8 ns, roughly corresponding to 16 kilodaltons, indicating that it is, indeed, a monomer in the experimental conditions used.

Validating inhibitor binding by protein-based NMR. The fingerprint spectrum for PRL3 is shown in **Figure 4.7**. As discussed in previous chapters, these resonances are sensitive to their environment such that chemical shift perturbations can be used to directly interrogate the binding of a protein to other molecules, such as, in this case, potential inhibitors. Protein-based NMR was therefore used to validate the binding of some experimental inhibitors to PRL3. Three experimental inhibitors were validated: compound 43, thienopyridone, and pentamidine.

As expected, compound 43 did not induce any CSP in the fingerprint spectrum of PRL3, indicating that it does not interact with it (**Figure 5.7**). Compound 43 is an inhibitor of PRL1 and acts by disrupting the trimer interface and preventing trimerization (*292*). As evinced by SEC and TRACT data, PRL3 is a monomer and is thus not expected to be inhibited by a modulator of trimer formation.

Figure 5.7. Characterizing PRL3 binding of compound 43 and pentamidine.

The fingerprint spectrum of apo PRL3 is shown in black and that of compound 43 in green (A) and pentamidine in blue (B). It is evident that there is no chemical shift perturbation in the presence of these compounds.





Meanwhile, pentamidine did not induce CSP in PRL3, even at three-fold excess. This indicates that pentamidine does not bind to PRL3 (**Figure 5.7**). The pentamidine formulation used in literature is the standard therapeutic grade Pentam 300 (296). This particular formulation is pentamidine isethionate. Interestingly, isethionate has a sulfonate (SO₃-) functional group that might mimic a phosphate group, which is known to favor the closed form of PRL3 (278). In the present study, isethionate was not present in the pentamidine (APExBIO, TX). *In vitro* inhibition assay using DiFMUP as a substrate also indicate that pentamidine alone does not inhibit PRL3.

Thienopyridone, meanwhile, induced significant CSPs across the entire spectrum of PRL3 (**Figure 5.8**). DiFMUP assays also indicate that it is able to inhibit the phosphatase activity of PRL3. However, thienopyridone and the more potent analog, iminothienopyridinedione (ITP), were shown to non-specifically inhibit protein phosphatases via oxidation of the catalytic cysteine. In the case of PRL3, this oxidative mechanism favors the formation of a disulfide bond between C49 and the catalytic C104 such that the C49A mutant is not inhibited by either compound in *in vitro* assays (297). Addition of up to four-fold excess of thienopyridone to ¹⁵N-labelled PRL3 harboring a C49A mutation does not induce any CSP in the fingerprint spectrum further confirming the non-specific mechanism of action of thienopyridone (**Figure 5.9**).

Figure 5.8. NMR on the interaction of PRL3 and thienopyridone. The fingerprint spectrum of apo PRL3 (black) overlaid with PRL3 in the presence of thienopyridone (red) indicate that thienopyridone alters the chemical environment of PRL3.



Figure 5.9. NMR on the interaction of PRL3.C49A and thienopyridone. The fingerprint spectrum of the C49A mutant of apo PRL3 (black) overlaid with the spectrum in the presence of thienopyridone (red) indicates that thienopyridone does not affect the C49A mutant. This is in line with the evidence that thienopyridone inhibits PRL3 by altering its redox state.



The current results indicate that proposed inhibitors must be validated specifically to ensure that their mechanisms-of-action are through direct interaction with their protein targets. Moreover, the results further illustrate the power and utility of NMR spectroscopy as a tool, not only for drug discovery, but for characterization of the interaction of proteins and potential inhibitors. In the case of PRL3, several other proposed inhibitors must be validated, and protein-based NMR is a viable strategy assuming these compounds are soluble.

Meanwhile, that some of the most potent PRL3 inhibitors such as pentamidine, thienopyridone, and ITP are, in fact, not active binders and inhibitors indicate that there remains a critical need to identify other potential inhibitors with a specific mechanism-of-action, involving specific binding to PRL3.

Drug repositioning screen identifies candesartan and salirasib as PRL3 inhibitors. One strategy to finding small molecules that modulate a protein target is drug repurposing or repositioning. This strategy involves assigning a new use – either as a new indication or as modulator for a new target – for a drug that already has an approved indication (*298–301*). Compared to the discovery of a completely new molecule, drug repurposing has a relatively lower risk and requires less investment as the drug has already been found sufficient and approved for human use (**Figure 5.10**). In most cases, the approved drugs already have desirable properties and with known mechanisms-of-action for their original target. It is worth-noting though that several FDA-approved drugs do not actually have wellcharacterized mechanisms-of-action as this is not a requirement for FDA approval

(302). While several drugs have been assigned new indications, a systematic approach to drug repositioning has only begun to emerge in the recent decade as a viable strategy for drug discovery (299, 303). In this work, a panel of FDA-approved drugs was screened for their ability to inhibit PRL3 and PRL3-mediated metastasis (304). Cellular assays discussed in this work were done in the laboratory of Jessica Blackburn at the University of Kentucky.

Figure 5.10. Risk and reward of drug discovery strategies. Drug repositioning or repurposing of already approved drugs (red box) provides an opportunity for drug discovery that is relatively low risk but with a potential for high reward compared to other strategies. Another high reward approach is *de novo* drug discovery (blue box) that may involve screening curated libraries *in vitro* or large to ultra-large *in silico* screens.



Around 1400 compounds from the L1300 FDA-approved Drug Library (Selleck Chemicals, TX) were screened for their ability to inhibit the phosphatase activity of PRL3 using the assay described in Chapter 4. Compounds that interfered with this assay (53/1433) were excluded from the screen. The remaining 1380 compounds were assayed and were considered a hit if they were able to reduce the phosphatase activity of PRL3 to three standard deviations lower than the mean activity for the entire screen.

Twenty compounds were identified as potential inhibitors given the screening criteria. Inhibition by these compounds were validated and nine were chosen on the basis of their ability to inhibit PRL3 and their IC50 values (**Figure 5.11**). These compounds are classified as broad PRL inhibitors as they are also able to inhibit the phosphatase activities of PRL1 and PRL2 significantly. Of the original twenty hits, four compounds specifically inhibited PRL3 (not shown) but were not chosen to move forward based on higher IC50s compared to the nine broad PRL inhibitors.

To assess the effect of these compounds on cell viability, an MTT (3-[4,5dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) assay was performed (**Figure 5.12**) (305). Briefly, HEK293T cells (ATCC, VA) were seeded in a 96-well plate at a density of 2 x 10^4 cells/well, allowed to recover for 24 hours prior to drug treatment for 16 hours. Drugs were added at the determined IC50 against PRL3. Drugs that had minimal effect on cell viability were chosen from this panel. Excluding Vitamin B12, these are candesartan and salirasib.

Figure 5.11. Broad PRL inhibitors from the FDA drug screen. The nine broad PRL inhibitors identified in the screen were validated in triplicate. All showed significant inhibition (***, p < 0.0001) relative to the negative control, dexamethasone (A). Dose dependence of these validated hits allowed for the classification based on high (B) or low (C) IC50. This image was adopted from Figure 1 and 2 of *Rivas and dela Cerna, et al. (2021) Sci. Rep. 11:10302,* licensed under CC-BY-4.0. Image was cropped and panels rearranged, resized, and combined without further modification.



Figure 5.12. Salirasib and candesartan are non-toxic. The identified inhibitors were characterized in an MTT assay in HEK293T cells to be minimally toxic (A), to increase cell viability (B), and to be highly toxic (C). Candesartan and salirasib were chosen based on minimal toxicity at their IC50. This image was adopted from Figure 3 of *Rivas and dela Cerna, et al. (2021) Sci. Rep. 11:10302,* licensed under CC-BY-4.0. Image was cropped and panels rearranged without further modification.



As discussed above, PRL3 has been implicated in metastasis and its expression is correlated with increased cell migration (REF). A PRL3-directed inhibitor should then be able to reduce cell migration in cell lines overexpressing PRL3 (**Figure 5.13, 5.14**). To ascertain the ability of candesartan and salirasib to inhibit cell migration, a scratch assay was done (*306*). HEK293T cells transfected with a PRL3 expression vector and HCT-116 (ATCC, VA) and SW480 (ATCC, VA) colorectal cancer cell lines that have high endogenous PRL3 expression were used. Both salirasib and candesartan inhibited cell migration in these cells in a PRL3-dependent manner.

Figure 5.13. Inhibition of PRL3-induced HEK293T migration. HEK293T cells were transfected with PRL3 resulting in increased migration (A). This is reversed by treatment with salirasib (B) and candesartan (C) in a PRL3-dependent manner. *p = 0.089 using one tailed student t test, **p = 0.03, ***p = 0.001, ns = no significant difference, using a two-way ANOVA. This image was adopted from Figure 4 of *Rivas and dela Cerna, et al. (2021) Sci. Rep. 11:10302,* licensed under CC-BY-4.0. Image was used without further modification.



Figure 5.14. Inhibition of migration of colorectal cancer cells. Representative images from scratch assay in HCT116 showing the inhibitory effect of candesartan and salirasib (A). Results from scratch assay from triplicate experiments in HCT116 and SW480 show that candesartan and salirasib have significant effect on cell migration (B). The effect is also dependent on the expression of PRL3 as confirmed by knockdown of PRL3 (C). This image was adopted from Figure 5 of *Rivas and dela Cerna, et al. (2021) Sci. Rep. 11:10302,* licensed under CC-BY-4.0. Image was used without further modification.



Salirasib or farnesylthiosalicylic acid (**Figure 5.15**) is an anti-cancer drug that dislodges all Ras isoforms from the plasma membrane (*307*, *308*). These Ras isoforms contain a CAAX box that are targeted by farnesyl transferases allowing them to be farnesylated and be localized in the plasma membrane (*309*). Interestingly, PRL3 is also farnesylated and localizes to the plasma membrane in a similar mechanism as Ras proteins (*310*, *311*). It is, then, possible that salirasib also affects PRL3 function inside the cell by altering its localization. It is worth noting that while salirasib inhibits cell migration in HEK293T cells overexpressing PRL3 and in both colorectal cancer cell lines, further reduction is observed even when PRL3 is knocked down indicating 'off-target' effects. Although, technically, as this is a repositioning screen, perhaps its action on PRL3 is the off-target effect. If the utility of salirasib as a PRL3-directed inhibitor is to be pursued, its interactions with other cellular pathways must certainly be taken into consideration.

Meanwhile, candesartan (**Figure 5.15**) – or related formulation, candesartan cilexetil which is ultimately metabolized to candesartan – is an angiotensin II type 1 (AT1) receptor blocker that is approved to treat hypertension or high blood pressure (*312*, *313*). Unlike salirasib, the effects of candesartan on cell migration appear to be PRL3-dependent indicating a mechanism-of-action that is perhaps more specific. This difference also highlights an interesting observation that anti-cancer drugs, like salirasib, are less likely to be successfully repositioned (*301*).

Figure 5.15. Salirasib and candesartan. Salirasib (left) is an anti-cancer drug that targets the localization of Ras proteins while candesartan (right) is a heart-failure medicine that target the angiotensin receptor.





Given their ability to inhibit the phosphatase activity of PRL3 and behavior in cellular assays, salirasib and candesartan both present as potential PRL3, or possibly broad PRL family, inhibitors but warrant further investigations on their cellular actions. That said, the PRL3-dependent action of candesartan in cell migration assays indicate, perhaps, a better potential to be repositioned as an anti-cancer PRL3-directed therapy.

Molecular mechanism-of-action by salirasib and candesartan. To further characterize how salirasib and candesartan interact with PRL3, Lineweaver-Burke analyses were performed, extracting initial velocities from kinetics studies using similar conditions as used in the *in vitro* DiFMUP assays but using increasing concentrations of the substrate (**Figure 5.16**). Analysis reveals that for both candesartan and salirasib, the apparent V_{max} is reduced but the K_M is not affected. This indicates that both are non-competitive inhibitors of PRL3.

Figure 5.16. Lineweaver-Burke analysis of candesartan and salirasib. Both salirasib (A) and candesartan (B) are non-competitive inhibitors based on their reduced V_{max} associated with no change in K_M . Right side of each panel are zoomed in on the boxed areas on the plots. All experiments ran as triplicates. This image was adopted from Figure 6 of *Rivas and dela Cerna, et al. (2021) Sci. Rep. 11:10302,* licensed under CC-BY-4.0. Image was cropped without further modification.



To show that these inhibitors directly bind PRL3, protein-observed NMR was attempted. Unfortunately, both candesartan and salirasib are not soluble in aqueous solutions – candesartan has a Log(P) of 6.10 and salirasib has a Log(P) of 6.8. Log(P) is a measure of hydrophobicity of a compound, where P is the octanol-water partition coefficient. A high Log(P) indicates that a compound is more soluble in octanol (hydrophobic) than water (hydrophilic). As such, to generate a hypothesis on the PRL3-drug interactions, molecular docking simulations were instead done (**Figure 5.17**).

Since there is no *a priori* knowledge on the conformation that the compounds bind to, two structures representing the open and closed conformations of PRL3 were selected as receptors for the *in silico* docking (278, 279). Autodock4 was used to dock salirasib and candesartan blindly on both the open and closed conformations of PRL3 (314). In this blind docking approach, the center of the grid box was aligned with the center of the molecule and the dimensions were set to maximum to cover the entirety of PRL3. For each ligand, 100 dockings were performed and clustered using an RMSD tolerance of 2 Angstroms. The most probable pose was selected based on calculated binding energy and the population of the cluster.
Figure 5.17. Molecular docking of salirasib and candesartan. The twenty FDAapproved drugs identified in the current study bound to one of two sites. Site 1 is adjacent to the WDP loop while Site 2 is adjacent to the P-loop (A). Salirasib binds to Site 1 and fits into a hydrophobic pocket near the C-terminal while candesartan binds to a potentially new pocket, Site 2 (B).





В

Α

Site 2



Modeling suggests that salirasib binds (free energy of -10.46 kcal/mol) to a pocket in the closed conformation of PRL3, adjacent to its WPD loop and close to a hydrophobic pocket that has been proposed to fit a farnesyl chain (279). The farnesyl half of salirasib indeed fits in this pocket, with the salicylic acid group facing the outside of the protein and hydrogen bonding with the backbone amide and carbonyl groups of G73, likely contributing to the stabilization of PRL3 in the closed conformation (Figure 5.17B). Candesartan, meanwhile, binds on the other side of this pocket, also in the closed conformation of PRL3. It binds (free energy of -10.26 kcal/mol) in a pocket adjacent to the CX5R loop and interacts with the side chain of K136 and backbone carbonyl of A106 (Figure 5.17B). Notably, the interactions do not involve any active site residues, indicating that these two sites are potentially allosteric sites and these two drugs, although with very minimal evidence at the moment, are also possibly allosteric. The WPD-adjacent site has previously been proposed to be the binding site of thienopyridone and ITP, although NMR studies in the present work do not indicate any interaction specific with PRL3 (264, 287). The CX₅R pocket, meanwhile, has not been previously pursued.

Hydrophobicity of these molecules limited the type of biophysical experiments that can be done, but current analyses maybe have uncovered potential druggable sites on PRL3. These pockets are worth pursuing in the future towards identification of PRL3-directed inhibitors.

Examining candesartan analogs. Candesartan binding to a potentially novel druggable pocket in PRL3 certainly warrants further investigation. However, solubility issues encountered with the drug prevents more detailed biophysical characterization of this interaction. A viable strategy is to look at related molecules and analogs that might be more soluble than candesartan, followed by structure-activity relationship to optimize the compound.

Thus, compounds structurally similar to candesartan were mined from PubChem (*315*, *316*). A total of 796 compounds were identified and docked to the closed conformation of PRL3 using an identical protocol used above. Of these, 34 docked with better docking scores than candesartan (free energy of -10.3 kcal/mol or better) with an average Log(P) or predicted Log(P) of 5.9, ranging from as low as 3.10 to as high as 9.8 (**Table 8**). While the current study did not further examine this avenue, this provides a clear and quick next step to investigating potential of candesartan-like molecules to inhibit PRL3.

	Binding Energy		
Compound CID	(kcal/mol)	LOG(P)	
142494141	-11.2	5.4	
22596772	-11.0	5.9	
10345537	-10.9	3.9	
19792192	-10.7	5.5	
19792152	-10.5	4.6	
54423843	-10.5	4.8	
15662012	-10.4	5.1	
19792177	-10.4	5.9	
69945758	-10.4	5.8	
53845115	-10.4	4.4	
70324433	-10.4	3.1	
68755817	-10.3	5	
15654763	-10.3	4.7	
15654738	-10.3	3.8	
141290740	-10.3	5.4	
140992535	-10.3	4.6	
142494141	-11.2	5.4	
22596772	-11.0	5.9	
10345537	-10.9	3.9	
19792192	-10.7	5.5	

Table 8. Candesartan derivatives (Top 15 with logP < 6) from *in silico* screening.

Towards a virtual screening approach targeting PRL3. With available computational resources, screening of large libraries is now readily available (*317*). Ultra-large virtual screening has been applied to identify inhibitors against the SARS-CoV-2 main protease from a library consisting of more than 250 million compounds (*318*). A screening campaign targeting KEAP1 and NRF2, meanwhile, assembled the largest ready-to-dock ligand library consisting of more than 1.4 billion make-on-demand compounds (*319*). These developments accelerate *de novo* drug discovery by making a large chemical space accessible. It is the intent of this section to merely spark inspiration to the reader who might be inclined to pursue this avenue for PRL3 (the author certainly will!).

Simply to illustrate a point, in this work, a very small subset of the Zinc library, specifically including compounds from DrugBank (n = ~10,000) were screened targeting both the open and closed conformation of PRL3 (*320, 321*). Virtual screening was done with following the same protocol as above (**Figure 5.18**). The closed conformation appears to be more amenable to binding ligands based on mean docking scores: mean free energy of -6.2 kcal/mol for open conformation against a mean free energy of -9.8 kcal/mol for the closed conformation. Based on this largely preliminary test virtual screen, it appears that the strategy might be able to yield potential binders and inhibitors. For the closed conformation, for instance, several ligands already have better docking scores than candesartan and other inhibitors previously identified.

Figure 5.18. Virtual docking of a small DrugBank library. A subset of the DrugBank library was docked blindly onto the open (blue bars, with grey line at the mean) and closed (red bars, with black line at the mean) conformation of PRL3. This preliminary docking study showed that some compounds with significantly higher docking scores than compounds already identified are accessible even with a small docking set.



Characterization of olsalazine, the first PRL3 binder. Thus far, no PRL3 inhibitor has been identified that has been confirmed to bind PRL3. Experimental inhibitors in the literature have not shown any confirmed binding to PRL3 using any biophysical tool like NMR or isothermal titration calorimetry (ITC), and the ones assessed in this work showed no observable binding. Meanwhile, inhibitors identified in this present work have not been amenable to biophysical studies due to solubility issues.

One of the goals of structural biology is to provide information that can be used, not only to understand protein function, but also to enable the design and optimization of inhibitors (*322*). There currently is no PRL3 structure bound to any inhibitors, limiting the opportunity for structure-based drug optimization. Thus, in an attempt to search for binders, remaining drugs from the FDA approved library that did not pass the screening criteria but show modest (>35%) inhibition were further screened by protein-based NMR. From this screen, there are 61 such drugs including the original 20 candidates discussed above. The 41 remaining drugs that showed modest (35-70%) inhibition of phosphatase activity relative to the DMSO control were subjected to protein-based NMR and the only criteria is the induction of chemical shift perturbation.

Out of the 41 screened drugs only one, olsalazine, showed CSP in the fingerprint spectrum of PRL3 (**Figure 5.19**). Addition of olsalazine to PRL3 resulted in significant chemical shift perturbation to residues within and outside the active site of PRL3 (**Figure 5.20**). NMR titration was done by adding 2 µl at a time of a

10 mM olsalazine stock into a sample containing 350 μ M of PRL3. The titration was done until no more additional CSPs are observed in the TROSY spectrum.

The backbone resonances of the PRL3:olsalazine complex is only partially assigned. Nonetheless, the CSPs were mapped onto the closed structure (PDB: 2MBC) of PRL3 (**Figure 5.20B**). CSP is observed across the entire structure. Interestingly, the highest CSP belongs to the phenylalanine residue in the WDP loop (red in **Figure 5.20B**). Whether olsalazine binds to the active site or modulates PRL3 activity by altering the conformation of the WPD loop is yet to be determined.

The inhibitory effect of olsalazine on PRL3 phosphatase activity is dosedependent with an IC50 of 200 μ M under the assay conditions (**Figure 5.21**). While certainly not the most potent inhibitor, it is the first ever molecule to be shown to inhibit *and* bind PRL3. Docking of olsalazine using the protocols outlined above was inconclusive and showed no preference for a specific site on PRL3. However, based on the fingerprint spectrum of the PRL3:olsalazine complex (**Figure 5.19 and 5.20**), the binding is specific – that is, only a single set of resonances are present indicating a single average structure for the complex.

Figure 5.19. Interaction of olsalazine and PRL3 by NMR. The fingerprint spectrum of apo PRL3 is shown (black). Chemical shift perturbation is observed upon the addition of small molecule drug, olsalazine (blue) indicating binding.



Figure 5.20. NMR titration of olsalazine onto PRL3. Titration of olsalazine into PRL3 revealed some fast exchange binding (marked with arrows). Chemical shift perturbation is seen in some active site residues (+) but also distal residues (A). The chemical shift perturbations are mapped on the structure and colored on a blue-to-red continuum with blue corresponding to no CSP (or unassigned) and red to highest CSP. The active site loops are labeled (B).



Figure 5.21. Dose-dependent inhibition of PRL3 by olsalazine. Inhibition of PRL3 by olsalazine showed dose-dependence with an average IC50 of 200 μ M. The individual trials (n = 4) are shown in different colors. Olsalazine was completely soluble in buffer and was used as a negative control (black circles). The structure of olsalazine is also shown.



To further confirm binding, olsalazine was titrated into PRL3 monitored by both NMR sand ITC. Titration by NMR reveal that some residues of PRL3 bind to olsalazine in the fast exchange regime, and these residues are both within and distal to the active site. For ITC, 1170 μ M of olsalazine was loaded onto the syringe and titrated into 52 μ M of PRL3. This revealed a binding affinity, K_D, of about 300 μ M, fitting the data to a one-site binding model (**Figure 4.11**). To date, olsalazine is the first and only inhibitor that has been validated to bind to PRL3.

Identification of olsalazine as a PRL3 binder opens up the possibility of determining the first ever structure of PRL3 in complex with a ligand/inhibitor. To begin PRL3:olsalazine structure determination, the backbone chemical shift resonances were partially assigned using a standard set of 3D experiments described in Chapter 3. There are 118 observable resonances in the fingerprint spectrum and 89 (75%) of these have been assigned.

Finally, the activity of olsalazine against other phosphatases was characterized using Phosphatase Profiler (Eurofins, LU). Out of 24 phosphatases included in this profiler kit summarized in **Table 9**, olsalazine only showed some inhibitory activity against 9 phosphatases – 55% inhibition, the highest, against DUSP3/TMPD while all the rest show less than 20% inhibition. This preliminary data indicates a degree of specificity for olsalazine, although this needs to be further validated against other phosphatases and kinases.

Phosphatase	Inhibition (%, n=3)
PP1	15
Lambda PP	_*
PP2A	-
PP5	6
LMPTP	-
MKP-5	2
VHR	-
PTEN	4
PTP1B	-
SHP2	11
PTPN22	-
PTPMEG	-
SH1	-
MEG2	18
PTPRB	19
CD45	-
RPTPM	18
TMDP	55
HePTP	-
TCPTP	-

Table 9. Specificity of olsalazine by phosphatase profiling

* (-) denotes no inhibition, some results showed negative inhibition

Olsalazine (**Figure 5.21**) is a small molecule consisting of two 5-aminosalicylic acids linked by an azo bond that is FDA-approved for the treatment of ulcerative colitis (*323, 324*). It is metabolized by gut bacteria, cleaving the azo bond, and liberating the active form, 5-aminosalicylic acid, in the large intestine (*325*). This bond cleavage is not expected under the experimental conditions *in vitro*, but it is interesting to point out that olsalazine is the second molecule identified in this work to have a salicylic acid group, salirasib being the other one. Whether this is an indication that salicylic acid-containing molecules can be used as scaffolds for PRL3 inhibitors, remain to be seen and might as well be a promising avenue to pursue.

PRL3 drug discovery, what next? The current work identified some molecules that might serve as possible starting points, or scaffolds, for further elucidation. A SAR-by-catalog approach might prove to be a promising approach for candesartan and its more soluble analogs, identified in the PubChem analog screen (*326*). Meanwhile, olsalazine and salirasib seem to point to the relevance of the 5-aminosalycylic acid group as a fragment that might be built upon in a fragment-based discovery approach (*163*). Overall, these molecules provide a new avenue of exploration to further accelerate PRL3-directed drug discovery.

The easily accessible and available chemical space – from pre-defined chemical libraries to massive make-on-demand collections – and the synergy between improvements in computational tools and virtual screening pipelines make the barrier to entry to a drug discovery program significantly lower. This work

demonstrates that beyond identifying potential inhibitors, it is very important to validate the binding and demonstrating a well-defined mechanism-of-action for candidate hits. NMR is a very powerful and highly informative tool for this purpose. In this work, only protein-based NMR was done, but ligand-observed NMR methods, some discussed in Chapter 3, are available solidifying its role in both the discovery of novel therapeutics and in the validation of previously identified compounds.

A few insights on PRL3 dynamics. Comparison of the open and closed forms of PRL3 reveal large-amplitude motions involving the active site loops, WPD and CX_5R (279, 327). The active site loops are about 18 Angstroms apart in the open conformation and are only 8 Angstroms apart when PRL3 is closed (**Figure 5.5**). Loop opening and closing has been observed in other phosphatases and has been well-characterized in YopH and PTP1B, for instance. In fact, the motions of these loops have been correlated to the rate of turnover for these phosphatases (*328*). Whether this specific dynamics-function relationship is generalizable to all phosphatases remains to be investigated.

An interesting observation is the degree of difference in the open and closed forms in phosphatases. In PTP1B, for example, the open/closed transition is characterized only by the movement of the WPD loop, very subtle compared to that of PRL3 (279, 328–330). In PRL3, both active site loops appear to move to arrive at the closed state and these active site loop motions are also accompanied by a twisting of the bottom half of the protein with respect to its top half, resulting

to the separation of strand 1 and helix 5 (**Figure 5.22**). This large-amplitude motion, in fact, contributes to the formation of the potentially druggable pocket adjacent to the CX_5R motif identified above. The significance of these motions as they relate to the phosphatase activity of PRL3, if any, remain unknown. It appears, however, that PRL3 exhibits significant flexibility – a possible reason why it has eluded structure determination by crystallography. The only structure of PRL3 determined by crystallography is in complex with the CBS domain of CNNM3, a magnesium transporter, which it interacts with as a pseudo-phosphatase (*327*).

Figure 5.22. Conformational changes upon open-to-closed transition. The conformational changes accompanying the open-to-closed transition in PRL3 is characterized by more than just loop motions as opposed to other phosphatases. Conformational changes in Site 1 (left) and Site 2 (right) are shown. Helix motions are marked with red arrows showing the transition from open (green) to the closed (grey) structure. (PDB: 1V3A, 2MBC)



In a preliminary attempt to further examine this flexibility, unbiased MD simulations were performed using three structures of PRL3: the closed (2MBC) and open (1V3A) forms used in the molecular docking experiments, and another open form (1R6H) (277-279). Simulations were performed on the University of Louisville's Cardinal Cluster using the CHARMM36 force field in GROMACS (331-333). Systems were subjected to energy minimization followed by a two-phase equilibration, first under an NVT (fixed number of atoms, N, volume, and temperature) ensemble, followed by equilibration under an NPT (fixed number of atoms, pressure, and temperature) ensemble. Each phase was done for 500 ps. After the equilibration steps, the temperature, pressure, and density were verified to be stable. The total production runs for each system are 102 ns, 70 ns, and 44 ns, respectively. To monitor the opening and closing of the active site loops, the distance between the C α carbons of active site residues C104 and D72 was used. To monitor the opening of the potentially druggable pocket, the distance between the C α carbons of residue V10 in the strand and Q135 in the helix was used. Looking at these two parameters reveal that PRL3 samples a wide range of conformations (Figure 5.23). Further analysis of these simulations, and longer simulation times, will provide more insight as to the flexibility and dynamics of PRL3. Some NMR methods that are suitable to further interrogate these dynamics properties of PRL3 were discussed in Chapter 3 and will be applied to GMPK in Chapter 7. The present simulations and preliminary analyses, nonetheless, provide some basic insight regarding the flexibility of PRL3 and also provides a pool of structures that may be analyzed with NMR data to be collected in the future.

Figure 5.23. Molecular dynamics simulation of PRL3. Active site loop motions and the opening of site 2 were monitored in the simulations of PRL3 using three different starting structures. Both parameters show a wide range of assumed values indicating possible large conformational changes that PRL3 can undergo.



Understanding the conformational heterogeneity of a system, as will be seen in the next chapter, and specifically identifying conformational states that are relevant to protein function may aid in design of effective inhibitors. Thus, it is important to also consider these dynamics in developing drugs against PRL3.

Functionally distal site mutations as tools to probe function and dynamics.

This chapter primarily focused on approaches to identify inhibitors of PRL3. Protein dynamics, and allostery, has been briefly discussed here and in previous chapters but will be covered more in subsequent chapters. Nonetheless, as mentioned above, an understanding of the structure-dynamics-function relationship in proteins can be leveraged to identifying effective inhibitors and possibly guide optimization of selective inhibitors, such for highly conserved domains like phosphatases (*334*, *335*). One approach to interrogate the dynamics of a protein is by looking at the effect of distal mutations that minimally perturb the protein structure but have an effect on function (*336–338*).

Four mutations that are away from the active site were chosen using the BioMuta database (**Figure 5.24**): two predicted to be benign (T32A and S143N) and two predicted to be damaging (R18H and P25H) to PRL3 function (*339*). Mutations were introduced to the expression plasmid using the Q5 mutagenesis kit (NEB, MA). Circular dichroism and NMR were used to verify that no significant structural changes are caused by the mutations (**Figure 5.25**). CD spectra for all four mutants are identical to the wild-type spectrum indicating no change in secondary structures. Similarly, fingerprint NMR spectrum are characteristic of

point-mutations and no significant CSPs are induced. These indicate that these mutations preserved the average structure of wild-type PRL3. Interestingly, none of these mutations reduced the phosphatase activity of PRL3 significantly but the P25H significantly increased its activity (**Figure 5.26**). In YopH and PTP1B, as mentioned above, motions of the active site loops are correlated with product turnover (*328*). Comparing the loop motions in P25H and the wild-type PRL3 by NMR relaxation dispersion, discussed in Chapter 3 and again in Chapter 7 is the first step to characterizing this possible correlation in PRL3. In enzymes, the chemical step is also only one of the processes that might be affected by dynamics. Binding of ligands and substrates to these mutants may also be regulated by dynamics. It would be interesting to see how substrate binding is affected, if so, in these mutants.

Figure 5.24. Some PRL3 mutations away from the active site. From the BioMuta database, four mutations in PRL3 were chosen that are distal from the active site loop. Two of the mutations are predicted to be benign to function (blue) while two are predicted to be damaging to PRL3 function (red).



Figure 5.25. Mutations do not affect the structure of PRL3. The four chosen mutations did not have any significant effect on the structure of PRL3. The fingerprint spectrum for the T32A mutation shows minor chemical shift perturbations characteristic of point mutations (A). Circular dichroism also reveals no significant changes in secondary structures (B).



Figure 5.26. Phosphatase activity of PRL3 mutants. DiFMUP assays reveal that the benign mutations (blue) do not have any effect on PRL3 activity. Meanwhile, the damaging mutant (red) R18H showed reduced activity. Interestingly, damaging mutant (red) P25H showed significant increase in phosphatase activity. All results are in triplicate and normalized to wild-type activity.



In this work, several approaches were employed to begin a drug discovery program targeting oncogenic and pro-metastatic phosphatase, PRL3. Specifically, virtual screens and library screens were used. Screening of an FDA-approved drug panel led to the identification of salirasib and candesartan as potent inhibitors of PRL3. Virtual screening of candesartan analogs has also identified potential inhibitors that have better properties in solution. Protein-based NMR screening allowed for the identification of olsalazine as the first PRL3 binder and the salicylic acid group as a potential fragment that may be used to develop inhibitors. Blind docking of a small DrugBank library, furthermore, indicates the possibility of identifying strong binders. Finally, this chapter ends with some insights on the flexibility of PRL3 and some approaches to study these dynamics. Subsequent chapters will focus on MDMX and GMPK, particularly relating to the applications of NMR to study structure and function, focusing in turn on applications of NMR beyond its power as a screening and validation tool in drug discovery.

CHAPTER 06

CONFORMATIONAL HETEROGENEITY OF THE p53-BINDING DOMAIN OF ONCOGENIC PROTEIN, MDMX

"Conformational heterogeneity is a defining characteristic of proteins."

Lyle and co. The Journal of Chemical Physics 2013 (340)

Conventional structure determination methods yield an average structure and fail to capture the conformational heterogeneity in biomolecules. NMR is a powerful technique that allows the interrogation of a protein conformational energy landscape and NMR observables can be used to create an ensemble of functional conformational states. In this chapter, NMR is applied to study the conformational heterogeneity of an oncogenic protein, MDMX, and provide some preliminary insight as to the challenges of targeting this important protein.

The guardian of the genome. p53 was originally discovered as a host protein bound to the SV40 large T antigen in virally transformed cells (*341*). While
originally classified as an oncogene, p53 was eventually identified as a tumor suppressor (*342*). Eventually, it was regarded as the 'guardian of the genome' based on the fact that its expression is triggered by DNA damage (*343–346*). In general, p53 functions as a sensor of cellular stress, and beyond DNA damage is also involved in sensing oncogene expression and ribosome dysfunction (*347–349*). In response to these cellular stresses, p53 activates the transcription of genes that regulate various stress response processes including cell cycle arrest, DNA repair, apoptosis, and senescence (*350–356*). The regulation and function of p53 is therefore critical in the development of therapeutic strategies against various cancers.

p53 and cancer. The TP53 gene that codes for p53 is the most mutated tumor suppressor gene in human cancers. In fact, almost 50% of all cancers involve a mutation in this gene (*357–359*). A majority of these mutations are missense mutations that abrogate the ability of p53 to suppress tumor growth and are mostly localized in the DNA-binding domain of p53 (*358*). These mutations involve either amino acids that participate in p53-DNA interactions or are mutations that weaken these interactions. Six such residues are considered hotspots for mutations as they account for roughly 30% of mutated residues in p53: DNA-interacting amino acids, R248 and R273, and amino acids R249, G245, R175, and R282 that affect DNA binding by altering the local conformations or the thermal stability of p53 (*357*, *358*). In addition to missense mutations that affect tumor suppressor function, some mutations also induce gain-of-function on p53 that ultimately transform it to

instead contribute to tumor growth and development (*359*, *360*). The prooncogenic functions of mutant p53 were first observed when mutant p53 transfected in TP53-null cells led to enhanced tumor formation in mice (*360–362*). This enhanced tumor formation and related oncogenic functions involve differential transcriptional regulation. To date, several genes have been identified that respond to mutant p53 and lead to enhanced cellular proliferation, evasion of apoptosis, chemoresistance, limitless replication, invasiveness, inflammation, angiogenesis, and others (*359*). Drug discovery efforts targeting p53, therefore, either attempt to reactivate the tumor suppressive functions or block pro-oncogenic functions of the mutant versions (*363*, *364*).

MDMX and MDM2: regulators of p53. Inactivation of p53 is one of the major mechanisms that allow cancer cells to be proliferative and to evade apoptosis. These phenotypes are typically caused by cellular stress such as DNA damage that cause cancers to become hyper-proliferative and are a consequence of an inactive p53 pathway (*359*, *365*). As mentioned above, in about 50% of human cancers, mutations in the TP53 gene results in missense mutations that either abrogate p53 tumor suppressive functions or result in gain-of-function mutations in p53 (*357*, *358*).

Meanwhile, in the other half of cancer cases that retain wild-type p53, its tumor suppressor function is regulated by inhibitory proteins such as MDMX and MDM2 (*353*, *366–370*). These two related proteins down-regulate the activity of wild-type p53 under normal conditions and their aberrant expression or genomic

amplification result in tumorigenesis (*371*). Mouse mdm2 was first identified as the product of gene amplification in acentromeric extrachromosomal bodies, also referred to as double minutes, in a transformed mouse cell line – hence, the name mouse double minute 2. The human homolog MDM2, also referred to as HDM2, was later identified, as well as another homologous protein MDM4/MDMX or HDM4/HDMX (*372*). One or both of these proteins maybe overexpressed in cancers and they may individually or in concert with each other inhibit p53 (*373–375*).

MDM2 and MDMX inhibit the tumor suppressor function of p53 by directly binding to it, particularly to p53's first transactivation domain (p53-TAD, **Figure 6.1**) (*374*). Both proteins have an N-terminal domain that binds to p53-TAD leading to its inhibition and reduced tumor suppressive functions in cancer cells. Meanwhile, MDM2 is also an E3 ubiquitin ligase that can tag p53 with ubiquitin directing it for protein degradation (*372*, *376*). In addition to their individual functions as p53 inhibitors, MDM2 and MDMX can collaborate with each other as a heterodimer (*373*, *377*). The ligase activity of MDM2 is enhanced by the formation of a homodimer or a heterodimer with MDMX leading to increased p53 degradation (*374*, *378*, *379*). With MDM2 and MDMX contributing to activation of oncogenic pathways on their own and in concert with each other, it is necessary to develop therapeutic strategies that inhibit both proteins simultaneously.

Figure 6.1. MDMX binds the first transactivation domain of p53. The structure of MDMX (light blue) bound to a peptide of the first transactivation domain of p53 (red, yellow mesh). The MDMX surface shown in mesh (grey) contains a pocket in which the peptide fits. (PDB: 2MWY)



There are significant similarities between MDM2 and MDMX, particularly in their p53-TAD-binding domains that, simply based on sequence and structure similarity, developing a dual-targeting inhibitor would be straightforward. The two proteins are 30% identical and 50% similar in their primary sequence. Both also contain an N-terminal SWIB/MDM2 domain, a zinc finger Ran-binding domain, and a RING/Ubox domain at their C-termini. Considerable parts of both are also intrinsically disordered. Focusing only on the SWIB/MDM2 domains, which contains the p53-TAD binding site, significantly increases the similarity between the two proteins. In their N-terminal domains, MDMX and MDM2 are 57% identical and 77% similar. While distinct in their activities, there is definitely significant conservation in the primary sequence between these homologous proteins, particularly in the p53binding residues (380). Despite these similarities, inhibitors directed at MDM2 have been ineffective at targeting MDMX (381–383). Understanding the mechanism of p53 binding is, then, paramount to uncovering the molecular basis of this enigma towards development of MDMX and dual targeting inhibitors.

Interacting with the genome's guardian. Homologous proteins, MDMX and MDM2, both interact with the first transactivation domain of p53 (Figure 6.1). There is significant similarity in the binding interface, resulting in a similar affinity of both proteins to the p53-TAD peptide (*380*, *383*). MDM2 binds the p53 with a dissociation constant, K_D , of 0.24 μ M, while the MDMX-p53 interaction has a K_D of 0.15 μ M (*383*, *384*). Furthermore, the binding mode of p53-TAD to MDMX and

MDM2 is very similar (*383*, *385*). Structures of these complexes have been determined revealing atomic-level details on the interactions (**Figure 6.2A**).

The MDM2 and MDMX p53-binding clefts are largely hydrophobic and predominantly interact p53 amino acids F19, W23, and L26, which are all positioned on the same side of the p53 helix facing the binding clefts (**Figure 6.2B**) (*385*). Notably, F19 and W23 are buried into the MDMX/MDM2 pocket, with the W23 residue of p53 having an identical position in both proteins (*383, 386*). The L26 residue of p53 resides in a pocket that does differ in both proteins. The MDMX pocket contains a methionine at residue 53, while MDM2 has a smaller leucine in corresponding residue 54 conferring a difference in the size of the L26 pocket (*385*). This pocket is also affected by Y99 in MDMX which may protrude into this L26 pocket in the closed conformation of MDMX, ultimately resulting into a smaller hydrophobic cleft compared to MDM2 (*387*). Overall, while there are subtle structural differences between these two domains, the general binding mode of p53 and the features of the p53 binding interface are conserved between the two proteins – enough so that their binding affinities are almost identical.

Figure 6.2. p53 peptide binding to MDMX/MDM2. The N-terminal/p53-binding domains of MDMX (blue) and MDM2 (purple) are very similar (A). The structure of the p53 peptide bound to MDMX (yellow) and MDM2 (salmon) with the p53 residues predominantly interacting with MDM2/MDMX shown as sticks. The binding modes are very similar, with some minor differences in the side chain orientation (B). (PDB: 2MWY, 4HFZ)



Nutlin blocks MDM2 but not MDMX? The degree of conservation in the binding pockets of MDMX and MDM2 and, certainly, their similar affinity to their endogenous binding partner, p53, or specifically the p53-TAD, may indicate that targeting one of these pockets will yield an inhibitor that is also effective in binding the other. The nutlin-3a scaffold was designed to mimic the p53 pharmacophore (**Figure 6.3**) and is a very potent inhibitor of the MDM2-p53 interaction, with a K_D of about 30 nM (*387*). Interestingly, nutlin-3a, as well as other compounds developed to target the MDM2-p53 interaction display significantly lower affinities for MDMX (*385*).

Figure 6.3. Nutlin was designed based on the p53 pharmacophore. The bound structures of MDM2 to p53 (cyan) and nutlin (green) are almost identical (A). Nutlin and p53 have a similar pharmacophore. The p53 peptide (gray) show that the residues that comprise its pharmacophore (red and labelled) match up with parts of nutlin. (PDB: 4HFZ, 4HG7)



Since the L26 pocket is also required for the binding of nutlin-3a, as a p53 pharmacophore mimic, one potential reason for this differential affinity is the composition of this pocket. Indeed, mutating M53 to a valine, increasing the size of the L26 pocket, leads to a five-fold improvement in affinity, from 25 μ M to 5 μ M. Meanwhile, a Y99T mutation does not improve this interaction but, in fact, results in an 8-fold reduction in affinity (*377*). This implies that the structure of the L26 pocket does impact the binding of nutlin-3a, even if it does not alter the binding mode of p53. However, regardless of this improvement, the binding of nutlin-3a to MDM2 is still more than 150-fold stronger than that with this improved MDMX.

In a study of a series of MDMX inhibitors, a potential role for protein dynamics was implicated in the difference in binding affinities of structurally related molecules and in the difference in affinity of MDMX and MDM2 towards nutlin (*383*). Unlike MDM2, there is evidence that the apo form MDMX exhibits dynamics on the intermediate to fast NMR timescales such that ~30% of backbone amide groups are not observed in the fingerprint spectrum. It is important to note that no crystal structure of the apo form of MDMX is available, as well, perhaps due to this inherent flexibility. Upon addition of p53, most of the missing resonances, largely localized near the L22 pocket, appear and the intrinsic backbone dynamics is dampened. Small molecules that similarly engage this pocket and dampen MDMX dynamics show increased affinity compared to those that do not (*383*).

Future MDMX-directed drug discovery efforts should therefore consider this L22 pocket for higher affinity inhibitors. Moreover, this study implicates dynamics in differential binding of small molecules to two structurally similar binding pockets.

Given the observed dynamics in MDMX, it is therefore crucial to understand its conformational heterogeneity. Such information can be leveraged to developing high-affinity inhibitors against MDMX.

Functional conformations vs average structure. This work aims to provide some preliminary insight on the conformational heterogeneity of MDMX to support the hypothesis that functional conformations of MDMX and MDM2 are distinct from their average structures. As discussed above, MDMX and MDM2 bind p53-TAD in a similar manner and with almost identical binding affinity. These 'average structures' are those that are obtained from conventional structure determination methods as discussed in Chapter 2. These structures typically correspond to a low-energy state that is highly populated in the protein's conformational energy landscape. In early discussions of energy landscapes and protein motions, the idea that some conformations are not necessarily important or coupled to functions have been explored (69). Now, there is emerging evidence that the conformational energy landscapes of proteins indeed contain conformational substates from which protein function emanates, to borrow language from the 'structure determines function' paradigm (*388*).

In the case of MDMX and MDM2, it is hypothesized that the average structure determined are different from the functional conformations, such that static structures cannot account for the differences in their binding affinities to the small molecule inhibitor, nutlin-3a (*385*). Furthermore, these functional states may well be sparsely populated, transiently sampled, undetectable by conventional

structural biology tools. To fully understand the role of these functional conformations in determining the ability of MDMX and MDM2 to interact with inhibitors, it is important to characterize, at the atomic level, these sparsely populated and transient conformational substates of both proteins.

Recap: NMR is a powerful tool for protein dynamics. Molecular function relies on conformational dynamics ranging from picoseconds to beyond hours (**Figure 3.2**). The ability of NMR to sample this wide range of timescales makes it a very powerful tool in understanding protein dynamics. Additionally, NMR is able to probe these timescales at atomic level and near physiological environments (77, *389–391*). Furthermore, observables from solution NMR directly report on protein motion and is thus the ideal tool to study functional dynamics of proteins.

Here, residual dipolar coupling (RDC) analysis is used to probe the conformational heterogeneity of MDMX. RDC analysis is able to characterize the amplitude of motions spanning a wide range of timescales of motion, including the inaccessible 'hidden time' window that occurs between the tumbling time and about 40 µs. As such, RDCs are an excellent tool to study protein dynamics.

Affinity of MDMX to nutlin-3a. To determine the binding affinity of nutlin-3a to MDMX and MDM2, ITC was used with MDMX/MDM2 in the cell and nutlin-3a as a titrant in the syringe (**Figure 6.4**). The calculated K_D for MDM2 is 15 nM, within the same order of magnitude as what is known in literature (*392*). Meanwhile, the K_D for MDMX was determined to be 3.8 μ M. This value is slightly lower than what is

known in literature (around 25 μ M), but the MDMX constructs used are different (377). The current constructs are minimal and only contain residues 7-111 of the p53-binding domain. That said, even with this dissociation constant, the affinity to MDMX is still almost 250-fold weaker than that with MDM2.

To further confirm the binding affinity for MDMX, an NMR titration was done (**Figure 6.5**). This yielded a K_D in the same order of magnitude as that determined from ITC, at 8 μ M (**Figure 6.6**). Based on current experiments and literature value, the affinity of nutlin-3a to MDM2 is in the low nanomolar range, while that of MDMX is in the low micromolar range. Based on these values, the affinity of MDMX to nutlin is indeed at least 250-fold (to almost 2000-fold, using literature values) weaker than MDM2.

Figure 6.4. Nutlin-3a binding to MDMX and MDM2 by ITC. Nutlin-3a binding is 250-fold stronger with MDM2 than with MDMX.



Figure 6.5. NMR titration of nutlin-3a into MDMX. Several peaks appear as nutlin-3a is titrated into MDMX indicating some ordering in MDMX. Previous studies have shown that MDMX is flexible in its apo form. Assignments are for the fully bound form (magenta). The apo spectrum is in brown, and titration points are in an orange-to-blue continuum.



Figure 6.6. Binding affinity from NMR titration. Fast exchanging peaks were fit to determine the binding affinity of nutlin-3a using the titration equation in Chapter 4. Maximum chemical shift difference was fit as a local variable while K_D was treated as a global variable. The final affinity was derived as an average of all the fits. Error bars are derived from Monte Carlo simulations



Small molecule binding to MDMX. There are a number of inhibitors that have been identified that bind to MDMX (*383*, *393*). While the affinities remain modest compared to the inhibitors of MDM2, these inhibitors – and the complex structures bound to MDMX – provide insight as to the potential role of dynamics in molecular recognition. These molecules all bind to MDMX with a similar pharmacophore as p53-TAD (*383*).

The complex structures of MDMX with these series of inhibitors show a conserved global fold (**Figure 6.7**). Yet, heterogeneity in the structures is apparent. Since the compounds share similar pharmacophores, the binding pocket of the p53-TAD was superimposed for all the structures using the p53 peptide-bound structure as a reference. Calculation of per-residue root mean square deviation (RMSD) reveals the heterogeneity in this ensemble of structures. Higher RMSD for a residue indicate that with the binding cleft as a reference, this residue differs significantly in its positioning in the three-dimensional space with respect to the residue in p53-bound MDMX (**Figure 6.7**).

Figure 6.7. Structures of MDMX reveal heterogeneity in distal sites. Several structures of MDMX (colored differently) bound to small molecule ligands with similar pharmacophore to p53 are available. Structures were superimposed based on their p53 binding sites and the RMSD was calculated for each residue. Increased RMSD is seen in helix 1, loop 1, and helix 3.



Significant differences are observed in several parts of MDMX outside the binding site. Several structural elements display large differences in the N-terminal region, helix 1, loop 1, and the C-terminal region of MDMX starting from helix 3. These structural elements are all distal from the binding pocket indicating that interaction of small molecules with the p53 binding pocket in MDMX are allosterically linked to these distal regions. Moreover, the flexibility of these distal regions might contribute to the ability of MDMX to bind to small molecules.

Conformational heterogeneity from existing structures. As discussed above, comparison of existing MDMX structures bound to p53 or small molecule inhibitors show some heterogeneity in regions distal to the binding pocket. To further analyze these structures, principal component analysis (PCA) was performed. PCA allows for the investigation of the relationship of the structures to conformational changes associated with the variability. The plot of mode 1 and mode 2 (**Figure 6.8**) confirm the RMSD analysis and shows that these structures are very distinct from each other. Some structures also appear closer to the p53-TAD. Interestingly, all X-ray structures form a single cluster despite being bound to different small molecules. Structures determined by NMR, even those that have similar pharmacophores such as the SJ series have more distinct structures.

Figure 6.8. Principal component analysis of existing MDMX structures. Plot of the two largest modes/principal components show a cluster of similar structures but also other that are unique but bound to the same series of inhibitors with the p53 pharmacophore. In this figure, each structure in the 20-member bundle NMR structures is shown as individual points. It is important to note, however, that these are equally valid structures.



The specific modes derived from PCA correspond to specific movements in the distal sites, with the first mode accounting for most of the variability in the structures. The regions identified above are involved in motions that are captured in the top modes. Specifically, the relative movement between loop 1 and helix 4 appear to be the most dominant motion accounting for the variability and heterogeneity in structures in small molecule bound-MDMX (**Figure 6.9**). This conformational heterogeneity and movement in these distal sites potentially contribute to the challenges associated with developing inhibitors against MDMX.

Figure 6.9. MDMX motions contributing to largest variance in PCA. The significant motion identified in PCA is characterized by the movement of loop 1 and helix 4 (labeled). The structures are colored based on conformational change amplitude: red with no significant movement, and white/blue with the highest change.



RDCs reveal that MDMX conformational heterogeneity is underestimated by structures/models. Existing structures of MDMX already reveal a degree of heterogeneity. However, as these structures are all average structures, residual dipolar couplings were measured to further characterize the conformational heterogeneity of MDMX.

Residual dipolar couplings (RDC) provide information on the orientation of the internuclear vector by two NMR active nuclei – such as N and H in an N-H bond – with respect to the static magnetic field. In the solution state, the time-dependence of the dipolar coupling interaction can be broken down into three components: overall molecular tumbling, internal motions, and bond vibrations. A molecule in solution does not assume a preferred orientation with respect to the magnetic field; it tumbles randomly. Since the dipolar coupling interaction depends the angle, θ , between the internuclear vector and B_0 , dipolar couplings are averaged to zero (129). In an environment that leads to partial alignment of the protein with respect to the magnetic field, the possible orientations for the internuclear vector are no longer populated equally, resulting to a non-zero residual dipolar coupling. The magnitude of the RDC is related to the time-averaged angle between the internuclear vector and the magnetic field (129, 131).

Thus, to measure RDCs, MDMX was placed in an anisotropic environment. Briefly, to a 1 mM ¹⁵N-labelled MDMX sample in buffer and in the presence of nutlin-3a, about 18 µg of Pf1 phage was introduced to induce alignment (*394*, *395*). In-phase Antiphase (IPAP)-type HSQC decoupled in the nitrogen dimension were

collected in both isotropic and anisotropic (phage) solution. The backbone amide RDC, ${}^{1}D_{NH}$, was calculated as follows:

$${}^{1}D_{NH} = (J+D)_{NH} - J_{NH}$$

The splitting in the nitrogen dimension under alignment media, $(J+D)_{NH}$, corresponds to the sum of the RDC and the scalar couplings, J_{NH} (Figure 6.10).

From the current experiment, measured at 600 MHz, 58 ${}^{1}D_{NH}$ were unambiguously collected, covering about 80% of the assigned MDMX backbone (**Figure 6.11**).

Figure 6.10. IPAP-HSQC analysis of MDMX. Residues F47 in isotropic and anisotropic (with phage) media. The RDC is calculated from the difference in the splitting in anisotropic media and in isotropic media.



Figure 6.11. Residual dipolar couplings in MDMX. The complete set of RDCs collected in MDMX. 58 $^{1}D_{NH}$ values were unambiguously collected with phage alignment.


Meanwhile, from the available structures of MDMX bound to a series of small molecules and p53-TAD, an 'average' structure was calculated as a representative of the experimental ensemble. This average structure was used to back-calculate RDCs. The root-mean square deviation (RMSD) between the backbone C α 's of the ensemble and the p53-bound MDMX structure (PDB: 2MWY) was calculated on a per-residue basis. Similarly, the differences between the experimental, nutlin-bound MDMX, RDCs and the RDCs back-calculated from the averaged structure were also calculated. These were then plotted against residue number. The RMSD from the structure shows a similar trend from the comparison of the individual structures, as expected. Meanwhile, the RMSD of the RDC qualitatively recapitulates the regions of variability in the PDB ensemble (**Figure 6.12**).

Figure 6.12. Comparison of RMSD based on structure and RDCs. An average structure derived from the average of the ensemble of individual structures was generated and the per-residue RMSDs relative to the p53-bound structure were calculated (blue). The RMSDs based on RDCs were also calculated. A similar trend, with distal regions showing high RMSD values, is observed in both analyses.



The agreement between the experimentally measured RDCs and backcalculated RDCs can be quantified by a quality score or Q-factor, defined as follows:

$$Q = \frac{\sqrt{\sum_{i} (D_i^{exp} - D_i^{calc})^2}}{\sqrt{\sum_{i} (D_i^{exp})^2}}$$

A lower Q-factor indicates a good fit to the experimental data. The quality scores from all structures used ranged from 0.26 for the best to 0.80. The best fit corresponds to the MDMX-WK298 complex, while the worst corresponds to the MDMX-SJ298 complex (**Figure 6.13**). The observed differences in the structural elements noted above are apparent in comparing the structures of these two complexes. Interestingly, SJ298 is the weakest among the SJ series of MDMX inhibitors (*383*). It is worth emphasizing that all these compounds share a pharmacophore, and it appears that the difference in the conformation in distal sites does affect how they bind to MDMX.

Figure 6.13. SJ298 and WK298 are the worst and best fit, respectively. Structures of the best (WK298, blue) and worst (SJ298, magenta) are overlaid (A). Back-calculated RDCs per residue are plotted (left) along with the corresponding Q-factor and comparison to experimental results (right) for both compounds (B).



Insights from RDC analysis of MDMX. Two major conclusions can be drawn from this current work. First, it is worth noting that the RDCs were collected on the MDMX-nutlin complex. The MDMX-WK298 complex, with a Q factor of 0.26, indicates that this complex might have a structure that is close to the nutlin complex, beyond simply having a similar pharmacophore – after all, all molecules complexed to MDMX available so far, also have a similar pharmacophore (**Figure 6.13**). More importantly, the result reveals that even with a preserved binding site or pharmacophore, the distal parts of the protein adopt different conformations, indicating that, in the case of MDMX, the entire protein structure has an impact on its interaction with small molecules (**Figure 6.12**). It is therefore important to consider the complete conformational breadth of a protein in understanding their functions and in rational drug design.

Ligand binding and conformational heterogeneity. The results in this work indicates that the conformational heterogeneity might be important to small molecule binding. Based on the data, specifically the motions detected distal to the binding site, it is apparent that these distal sites are linked to small molecule recognition. Thus, accessing the full conformational breadth of MDMX can be utilized to better understand how it binds small molecules.

So what does MDMX look like? NMR observables like RDCs provide an insight on the molecular motions, more specifically the amplitudes of bond vector motion,

at a wide range of timescales including the 'hidden time' window. Specifically, RDC captures the amplitude of the motions at these timescales. To fully comprehend the conformational energy landscape of MDMX, it is necessary to generate an ensemble of MDMX that captures the conformational heterogeneity in solution. This is possible using NMR observables like RDCs and others like Cross-Correlated Relaxation (CCR) rates, which provide complementary information on the amplitude of motion on the fluctuation between two inter-nuclear vectors from occurring within the picosecond to millisecond timescales (99). These observables can be used as constraints in MD simulations or can be used to generate an ensemble via selection from a large pool of structural snapshots. This experimentally guided ensemble will then be a more accurate representation of MDMX and can be useful in the development of more potent inhibitors and specific inhibitors that target the lowly populated, yet functionally relevant structural states that differ from the 'average' structures reported in the PDB.

Furthermore, this can be coupled with relaxation dispersion (RD) experiments, discussed in Chapter 3, which capture kinetic information, to study allosteric networking within MDMX (*396*, *397*). Preliminary data has been collected (by the Lee lab) using the E-CPMG method that identifies significant conformational exchange in various regions of MDMX, with all residues exchanging between at a rate of between 6,000 to 9,000 s⁻¹ (**Figure 6.14**) (*396*). Subsequent studies shall examine clustering of residues based on changes in exchange rates at different temperatures. With this data and coupled with an accurate ensemble, it will be possible to begin building an allosteric network within MDMX.

Figure 6.14. Preliminary high-power RD applied on MDMX. Three residues from various regions of MDMX (V27 (N-terminal), H54 (binding pocket), and L109 (helix 4)) show a similar time scale of conformational exchange, which indicate possible correlated motions. Moreover, the RD data show that the conformational exchange or k_{ex} in MDMX is not completely quenched by conventional RD and the high-power RD is necessary for studying exchange in MDMX.



In this chapter, MDMX is used as a model system to probe the inherent conformational heterogeneity of protein structures and how this can be completely missed by conventional structure determination methods. Furthermore, this work serves as a preliminary observation to the hypothesis that conformational heterogeneity plays a role in protein function – in particular, why it has been significantly more challenging to target MDMX than MDM2. To this point, it appears that MDMX has intrinsic flexibility that affects how it binds to small molecules. Here, NMR was used to examine the structures of a protein and how they relate to the conformational heterogeneity in solution. In the subsequent chapter, NMR will be applied to study the dynamics of GMPK. The methods used in this GMPK chapter are also applicable to MDMX and may be revisited by anyone interested in the future to uncover further the dynamic aspects of MDMX which could serve as a model protein to study the role of global protein motions in function.

CHAPTER 07

SOLUTION DYNAMICS OF

HUMAN GUANYLATE KINASE, GMPK

"Proteins are not static systems. They undergo a great variety of dynamic processes at ordinary temperatures."

Ringe and Petsko Progress in Biophysics and Molecular Biology 1985 (63)

As has been alluded to several times in the previous chapters, the true power of NMR is in its ability to probe dynamic properties of proteins in solution. In describing the catalytic mechanisms of enzymes, it is important to identify changes in the average molecular structure that accompanies the events leading to product formation. In addition, it is also critical to investigate how dynamics change concurrently. In this chapter, NMR is applied to probe the solution dynamics of human guanylate kinase and how these dynamics are affected by its interaction with a substrate, GMP, with the ultimate goal of understanding the dynamics associated with the full catalytic cycle of GMPK. The first application of high-power CPMG RD to an enzyme also appears in this chapter.

Nucleoside monophosphate kinases and their biological functions. Nucleoside monophosphate kinases (NMPK) are enzymes responsible for catalyzing the reversible phosphorylation of nucleoside monophosphates (NMP) to nucleoside diphosphate (NDP). NPMKs use adenosine triphosphate (ATP) as a phosphoryl donor and produce adenosine diphosphate (ADP) and the corresponding NDP as products. Biologically, NMPKs are critical enzymes in the synthesis of deoxyribonucleotides (*398*, *399*).

Deoxyribonucleotide synthesis can occur via two pathways: the *de novo* synthesis pathway and the salvage pathway (**Figure 7.1**) (*399, 400*). In the *de novo* synthesis pathways, nucleosides are synthesized from small molecule building blocks to ribonucleoside monophosphates. NMPKs use ATP to convert these to ribonucleoside diphosphates that are then acted upon by ribonucleotide reductases to form the corresponding deoxy form. Nucleoside diphosphate kinases (NDPKs) then generates the triphosphate form. Meanwhile, the salvage pathway utilizes intermediates from degradative pathways and begin by importing deoxyribonucleosides into the cell which are then phosphorylated by nucleoside kinases, NMPKs and NDPKs. Ribonucleotide triphosphates and deoxynucleoside triphosphates are the building blocks for nucleic acids, RNA and DNA, respectively. Thus, nucleoside kinases are enzymes essential for all life (*399*).

Figure 7.1. *De novo* pyrimidine and *de novo*/salvage purine synthesis pathways. Schematic for the pyrimidine *de novo* pathway (A) and the purine *de novo* and salvage pathways (B). The work focuses on GMPK which acts on GMP, where the purine *de novo* and salvage pathways converge in the synthesis of GTP/dGTP (lower right). This image was adopted from Figure 1 of *Villa, et al.* (2019) cancers 11:688, licensed under CC-BY-4.0. Image was used without modification.



Guanylate kinases in brief. Guanylate monophosphate kinases are a subset of NMPKs that specifically act on guanosine monophosphates. These kinases, referred to as GMPKs or GUKs, are conserved across kingdoms of life from bacteria to humans (*401–404*). Like other NMPKs, they catalyze a reversible phosphorylation of GMP to GDP, using ATP as a phosphoryl group donor (**Figure 7.2**) (*405*).

In humans, GMPK (GMPK) is the only known enzyme responsible for cellular GDP production, making it essential for cell survival (*183*). As with other GMPKs, it is involved in both *de novo* and salvage pathways for DNA synthesis (**Figure 7.3**). In addition to this essential biological function, GMPK also assists in the activation of anti-viral and anti-cancer prodrugs (*406*, *407*). Some pro-drugs that are activated by GMPK include 6-thioguanine, mercaptopurine, ganciclovir, and acyclovir, among others (*407–410*). Thus, GMPK has two-fold relevance to human biology: as an essential enzyme in the synthesis of biomolecules and in the activation of therapeutics.

Figure 7.2. Reaction cycle of human guanylate kinase. The catalytic cycle of GMPK follows a random sequential mechanism. It first binds the two substrates, ATP and GMP, in any order, prior to the catalytic step. This is then followed by the release of two substrates, ADP and GDP, in any order. This image was adopted from Figure 1b of *Khan, et al. (2019) J Biol Chem 294:11920-11933,* licensed under CC-BY-4.0. Image was cropped used without further modification.



Figure 7.3. Cellular functions of GMPK. Human guanylate kinase is the only GMP kinase known in humans. It is at the juncture of the *de novo* and salvage pathways leading to the production of dGTP/GTP. Additionally, it is involved in the activation of a number of pro-drugs such as ganciclovir, which requires phosphorylation to be activated within the cell. This image was adopted from Figure 2 of *Khan, et al. (2019) J Biol Chem 294:11920-11933,* licensed under CC-BY-4.0. Image was cropped used without further modification.



GMPK as a cancer target. Given its role in biosynthesis of DNA and RNA, GMPK is an essential enzyme for continued cell growth. Cancer cells, which exhibit increased cellular proliferation, are characterized by an altered metabolic state which includes elevated levels of GTP and dGTP (*411*, *412*). These biomolecules are not just building blocks of nucleic acids but are also an energy source for cellular processes like protein synthesis and used in cellular signaling mediated by GTP-binding proteins. As these processes are typically upregulated in actively proliferating cells, depletion of GTP levels is considered a valid therapeutic strategy against cancers (*413–415*).

Among enzymes involved in the biosynthesis of GTP/dGTP, IMP dehydrogenase (IMPDH) and guanosine monophosphate synthase (GMPS) are currently being targeted for cancer therapeutic purposes (*416–418*). IMPDH is involved in the early steps of GTP and ATP biosynthesis, particularly in the conversion of IMP to XMP in the *de novo* pathway, the first committed and rate-limiting step in the pathway (*419*). Meanwhile, GMPS converts XMP to GMP, immediately prior to the GMPK-catalyzed step (*420*). The depletion of GTP/dGTP levels by inhibiting these enzymes can be reversed by guanosine or guanine coming in from the salvage pathway (*420*). GMPK is at the junction of the *de novo* and salvage pathways, making it a potentially effective therapeutic target (**Figure 7.3**).

The utility of GMPK as a cancer target is still understudied. Recently, however, it was shown that knocking down GMPK leads to the reduction of cellular viability, proliferation, and clonogenic potential of lung adenocarcinoma cell lines. More

importantly, this does not affect the proliferation of human peripheral airway cells (*183*). This provides further evidence on the potential of GMPK to be a target in oncology drug discovery.

Structure of GMPK. The first structure of the human guanylate kinase was only recently determined and although the structures from several GMPKs of other organisms have been known for quite a while (*183*, *421*). Guanylate kinases have several conserved domains including the P-loop, the GMP-binding domain, and a LID region, with several residues being conserved across bacterial and eukaryotic GMPKs. The structure of GMPKs is also conserved across organisms, although some bacterial versions, like that of *E. coli* and *S. aureus* are multimeric instead of monomeric (*183*, *402*, *422*, *423*).

The solution structure of GMPK (**Figure 7.4**) was recently determined and indeed shows a generally conserved architecture as other GMPKs, including similar domain arrangements (*183*). This structure was also determined in the open form and adopts a similar U-shaped conformation that other GMPKs adopt in their open forms (*424, 425*). The CORE domain takes a central position in the three-dimensional structure linking the other two domains via hinge regions. The GMP-binding domain contains a pocket that fits GMP lined with charged patches. Similarly charged residues are found in the CORE/LID interface where the ATP-binding pocket resides. The GMP-BD and LID domains in this U-shaped, open conformation are oriented away from each other. Interestingly, while the human and mouse GMPKs only share 88% identity, the structures, including side chain

orientations, are preserved among the two, with differences mostly in the termini and residues in the hinge regions (*183*). While availability of this structure is a significant tool in drug discovery programs targeting GMPK, further structural studies are necessary to completely understand its catalytic mechanism. **Figure 7.4. Solution structure of GMPK.** The first structure of human guanylate kinase was determined in the open conformation and reveals a conserved NMPK architecture. The domains are identified as LID (yellow), GMP-BD (cyan), and CORE (magenta) (A). The NMR ensemble consisting of 20 of the lowest energy structures are also shown (B). This image was adopted from Figure 2 of *Khan, et al. (2019) J Biol Chem 294:11920-11933*, licensed under CC-BY-4.0. Image was used without modification.



Conformational change in GMPKs during catalysis. Several structures of guanylate kinase from other organisms in various bound states reveal that ligand binding induces conformational changes that lead to an open-to-closed transition (*425, 426*). While the closed conformation of GMPK is not yet known, small-angle X-ray scattering and analytical ultracentrifugation provide evidence for a similar transition, such that the liganded form of GMPK is more compact than the apo form (*183, 421, 427, 428*).

The recently developed structure prediction tool, AlphaFold2 (AF2), were used on GMPK and it reveals a closed conformation similar to the known closed structures of other GMPKs which also matches a homology model predicted using the mouse version (**Figure 7.5**) (57). It is perhaps expected that AF2 predicts the closed conformation given that the training set probably has more ligand-bound structures for GMPKs and related enzymes. Nonetheless, given the accuracy of AF2 on several other systems, this prediction might be useful in further understanding the conformational cycle of GMPK, while an experimental structure is lacking (*57*). At the very least, this model will aid in structure determination or in the design of experiments to further probe the conformation cycle of GMPK, such as some experiments that will be discussed below (*62*). **Figure 7.5. Open and closed conformations of GMPK**. The apo form of GMPK assumes an open conformation (A) while the substrate-bound form is closed (B, C). Substrate binding triggers a transition from the open to the closed conformation. The structure of the closed GMP/ADP-bound form of mGMPK has been determined (B) while the closed conformation of GMPK (C) was modelled using AlphaFold2. (PDB: 6NUI, 1LVG)



Chemical shift perturbation upon GMP binding. Resonances for backbone residues were assigned for GMPK in the nucleotide-free state as well as in the presence of excess GMP as discussed in Chapter 4. It is worth noting that while the solution structure of GMPK currently available is in the open form, it was determined in the presence of free phosphate (*183*). The magnitude of the effect of the presence of phosphate on the structure of GMPK is currently not known but there are significant chemical shift differences in the fingerprint spectrum of 'phosphate-bound' GMPK and the current apo form of GMPK. In the current investigations, a MOP-based buffer was used removing any potential interference from the phosphate groups.

Of the 188 non-proline residues, the backbone resonance for 175 residues in the apo-state and 165 residues in the GMP-bound state were successfully assigned (**Figure 4.9**). Addition of GMP induces chemical shift perturbation that affects GMPK globally as seen in the fingerprint spectra (**Figure 7.6**). Furthermore, titration revealed that binding of GMP to GMPK occurs in the slow exchange regime of NMR (**Figure 7.7**). Significant CSPs are found within and close to the GMP-binding domain including residues that are expected to bind GMP based on available GMP-bound structures (**Figure 7.8A**) (*426*). This is clearly seen when the CSPs are mapped onto the structure of GMPK (**Figure 7.8B**). **Figure 7.6.** ¹**H**, ¹⁵**N-HSQC of GMP-bound GMPK.** The fingerprint spectrum of apo GMPK (black) is overlaid with the GMP-bound form (red). Chemical shift perturbation is seen across the spectrum indicating binding. Assignments correspond to resonances in the apo form.



Figure 7.7. NMR titration of GMP onto GMPK. Representative titration points from apo (black) to GMP-bound (red) are shown. Evidence of slow exchange is seen across the spectrum (A). The red box corresponds to the chemical shift perturbation of residue E48 which is in mixed slow/fast exchange. The black peak (apo) disappears as the red peak (4:1 GMP:GMPK) appears. In the middle of the titration, both peaks are, in fact, visible (blue, 1.5:1; green, 2.5:1) in the spectrum. (B).



Α

Figure 7.8. GMP binding residues and GMP-BD show highest CSP. Chemical shift perturbations between nucleotide-free and GMP-bound GMPK with red bars indicating high CSP (one SD above the mean) and orange indicating moderate CSP (above the mean). Red broken line indicates mean + SD and orange broken line marks the mean. GMP-binding residues (S37, R41, R44, Y53, E72, Y81, T83, D101, D103, R148) are marked with black asterisks (*) within the GMP binding domain (GMP-BD), along with the residues adjacent the hinge region (unlabeled gray box) connecting the GMP-BD to the CORE domain (A). Chemical shift perturbations mapped onto the structure of GMPK with high (red) and moderate (orange) CSP colored accordingly (B).




GMP-binding residues S37, R44, Y53, D103, and R148 are among the residues that show significant CSP that are one standard deviation above the mean CSP across the structure (**Figure 7.8**). Residues near the hinge region connecting the GMP-BD and the CORE domain also exhibit significant CSP. As chemical shifts are very sensitive to their environment, it is not surprising to see the residues involved in GMP binding to show significant perturbation. The induced closing of GMPK and the movement of the GMP-BD, also perhaps rationalizes the increased CSP for this specific hinge region.

Overall, chemical shift perturbations seen across most residues, including most significantly in the GMP-BD and the GMP-BD adjacent hinge region, indicates changes in the environment of the residues of GMPK which most likely correlates with the open-to-close transition occurring upon GMP binding. As discussed above, while there is currently no structure of closed GMPK, such closing is expected and supported by other methods (*183*, *427*).

Backbone dynamics of GMPK. To characterize the backbone dynamics of GMPK, experimental ¹⁵N longitudinal (R1) and transverse (R2) relaxation rates, ¹⁵N-{¹H} steady-state NOE (hetNOE), and transverse cross-correlated DD/CSA relaxation rates (η_{xy}) were acquired at 14.1 T.

N- and C-termini residues of proteins are typically expected to be flexible. The flexibility of the termini of GMPK is captured in these experiments as decreases in the R₂, hetNOE, and η_{xy} and an increase in R₁ (**Figure 7.9**). This is evident in both the apo and bound forms. Furthermore, there is a notable increase in the flexibility

in the loop connecting helix 5 and helix 6 as evinced by a dip in the hetNOE in this region (**Figure 7.9C**). While this loop appears to have flexibility whether or not GMP is present, the flexibility is more pronounced in the apo form. The flexibility of this loop has been noted in previous work but whether this flexibility has a functional consequence is yet to be examined. The GMP-binding domain also shows some reduced flexibility in the fast timescale upon binding of GMP.

Figure 7.9. NMR relaxation parameters for GMPK. Experimental ¹⁵N longitudinal (A) and transverse relaxation rates (B), {¹H}-¹⁵N hetNOE (C), and ¹H-¹⁵N dipolar cross-correlation rate (D) for residues of GMPK in the absence (grey) or presence (red) of excess GMP at 298 K and 14.1 T. Secondary structures of GMPK are indicated above the plots and shaded in lime green (β strand) and sky blue (α helix).



To quantify backbone dynamics, model-free analysis, discussed in Chapter 4, was performed. The observables were first analyzed using three diffusion tensor models of increasing complexity: isotropic, axially symmetric, and fully anisotropic. The best models were selected using an F-test which determined a full anisotropic model for the open, apo form and an axially symmetric model for the GMP-bound, closed form. Based on the R_2/R_1 ratio, the τ_c for the open form is 14.3 ns while that of the closed form is slightly lower at 13.8 ns. The order parameters were than calculated from five different models (**Table 4**) and the best fit was determined by model selection using the Akaike Information Criterion (AIC). As discussed in Chapter 4, the order parameters describe site specific motion of the N-H bond vector on the picosecond-nanosecond timescale (*116, 429*).

It appears that GMPK is well-ordered with an average generalized order parameter, S^2 , for the nucleotide-free state of 0.86 (**Figure 7.10A**). Addition of the substrate, GMP, does not further increase the order parameter. In fact, the GMPbound state has a slightly lower S^2 of 0.84 (**Figure 7.10A**). Secondary structure elements have a slightly higher average S^2 of 0.89 in both states. The decrease in order parameters is more apparent in the C-terminal half of the protein (**Figure 7.10A**, **B**). This region of increased flexibility is highlighted in **Figure 7.10C** and corresponds to a huge portion of the LID domain which is involved in binding ATP.

The GMP-binding domain, however, does show increased ordering at the ps-ns timescale. This is an expected response to substrate binding. Specifically, all residues that are expected to directly interact with GMP showed increase in S^2 (**Figure 7.10C**).

Figure 7.10. Order parameter of GMPK and response to GMP binding. Lipari-Szabo order parameter, S², derived from experimental relaxation data by Model-Free Analysis (A) for GMPK in the presence (red) and absence (grey) of GMP. Difference in order parameters, Δ S², between the bound and apo form (S²_{gmp} -S²_{apo}) colored by either increased (magenta) or decreased (orange) order by residue (B). Secondary structures of GMPK are indicated above the plots and shaded in lime green (β strand) and sky blue (α helix). Residues involved in GMPbinding are highlighted in red circles and shown as sticks in the model for GMPbound GMPK in the closed form based on the mouse model (PDB: 1LVG) (left, C) while residues in the LID/ATP-binding domain corresponding to areas showing a general reduction in order parameters upon GMP binding are highlighted in purple (right, C).



In addition to the backbone dynamics, some residues were fit to models that were indicative of conformational exchange. This is further examined below.

Conformational exchange in the microsecond-millisecond timescale. The binding of substrates in guanylate kinases induces a large-amplitude motion characterized by an open-to-closed transition (*424*, *427*, *430*). These motions may fall within the microsecond-millisecond timescale and can be studied by relaxation dispersion (*77*, *122*, *431*, *432*). Since GMPK is expected to exhibit the same response to GMP binding based on previous data, motion in this time regime was characterized using high-power CPMG relaxation dispersion (*117*, *183*, *396*, *427*).

A ²H, ¹⁵N-labelled GMPK sample was concentrated to 900 μ M and was used in the CPMG experiment. For the GMP-bound state, three-fold excess GMP, ~2.7 mM was added to the sample. GMPK residues did not exhibit significant dispersion at 298 K and the $R_{2,eff}$ values were very small, implying that the exchange rates are outside the observable rates for the experiment at this temperature. Previous work on guanylate kinases and adenylate kinase have shown that the open-toclose transition at 298 K occur too fast (*433*, *434*). Thus, temperature was lowered to 277 K to capture these motions. Assignments at 277 K was obtained by collecting fingerprint spectra for both the apo and bound states at 298 K, 288 K, 280 K, and 277 K, following chemical shifts responding to the temperature change (not shown). Residues that were unambiguous were excluded from the analysis. All available data was then analyzed using ShereKhan (*146*). Residues that exhibited a drop in $R_{2, eff}$ of more than 2 s⁻¹ were considered to show dispersion

and are the only ones used to calculate the exchange rate, k_{ex} . Data was fit under the fast exchange regime and the Luz-Meiboom model (*146*, *435*).

After identifying residues that displayed dispersion, the data was subjected to a clustering algorithm to identify residues that can be fit to a single exchange rate. In both the apo and bound states, the residues were all fit to a single global exchange rate, based on Akaike Information Criterion (AIC), comparing the global fit to clustered and individual fits.

Interestingly, GMP binding does not quench the motion (**Figure 7.11**). As mentioned above, more residues are observed in the GMP bound state to have dispersion. For the apo state, the k_{ex} is 10, 800 ± 800 Hz, while for the GMP-bound state is 10, 600 ± 600 Hz. These exchange rates are not within the limits of conventional CPMG, for which the limit for the CPMG pulse frequency is usually between 1.5 - 2.0 kHz. In these experiments, high-power CPMG extended this pulsing limit to 6.0 kHz, allowing an accurate determination of the conformational exchange rates for GMPK. A representative set of dispersion profiles are shown (**Figure 7.11B**) for cases that are exhibiting dispersion in the both forms, in GMP only, or in apo only. The grey vertical line in the dispersion profiles indicate the limit of conventional CPMG and shows the improvement that is provided by using high power. As of this writing, this is the first application of high-power RD to study the conformational dynamics of an enzyme.

Figure 7.11. Microsecond-millisecond motion detected in GMPK by highpower CPMG. High-power CPMG relaxation dispersion experiment captures µsms motion across the entire structure of GMPK. Residues are colored based on the condition they show dispersion: both nucleotide-free and GMP-bound states (yellow), only in nucleotide-free state (blue), and only in GMP-bound state (red) (A). Representative dispersion curves for each set of residues for the nucleotidefree (black) and GMP-bound (red) states. The lines were fit using ShereKhan, Grey line at 2 kHz represents the limit of conventional CPMG experiments (B).



Among the assigned residues with detectable dispersion, several are located within the GMP-binding domain and near the GMP-BD/LID hinge. In some cases, active site dynamics induced by substrate binding may contribute to the ability of an enzyme to sample conformations optimal for substrate processing (Figure **7.11B**) (436, 437). It is notable that more residues exhibit dispersion upon addition of GMP which may indicate a similar phenomenon where in conformation optimization occurs in preparation for substrate processing. Coupled with data from MFA, there is some evidence of increased flexibility in the ATP binding site, which might serve to prime the GMP-bound complex for ATP binding prior to the chemical step. It also may be the case that conformational exchange is only quenched once the two substrates are bound. It will therefore be interesting to see how these dynamics are altered in the ternary complex, such as in the presence of ATP analogs, to mimic a pre-processing state, or ADP. Similarly, it is interesting to see how binding on the ATP-binding site in the absence of GMP affects these dynamics. These studies may provide some fundamental dynamics information as a model for conformational dynamics in ternary complexes. Potentially, it is possible that the conformational exchange in this timescale is only quenched when both GMP and ATP are bound to GMPK so as GMPK is not trapped in a state where only one of them is bound.

Overall, analysis of relaxation dispersion for the apo and GMP-bound states of GMPK reveal interesting features that warrant further investigation, particularly in painting a complete dynamical picture of the enzymatic activity of GMPK and the accompanying conformational cycle.

Some notes on Helix 3. Helix 3, which is located immediately after the GMPbinding domain, has been of special interest to understanding the dynamics associated with GMPK function. Specifically, crystallographic studies of the GMPbound yeast GMPK revealed elevated B-factors in helix 3, compared to the apo form (428). Van der Waals contact in this region was not disrupted and is not thought to be the reason for the increased mobility based on B-factors. As discussed above, binding of GMP induces a transition into the closed form that brings the GMP-binding domain closer to the LID and CORE domains (427, 428, 430). It was proposed that this motion acts to unwind the helix, causing a tension that increases mobility (Figure 7.12). This mobility is thought to assist in product release (428). The same hypothesis has been put forth in adenylate kinase (AK) where the Ap5A complex resulted in increased mobility in the analogous area, after the NMP binding site (438). In the apo form of AK in this study, mobility is observed in the NMP-binding domain which is quenched upon substrate binding – or as proposed, mobility is moved outside the NMP binding domain after successful substrate binding. It was thought that this flexibility provided a counterweight to ensure product release. That is, so that the complex is not trapped in an energy well (438). Interestingly, in the case of mouse GMPK, this increase in B-factor is not observed in the ternary complex with ADP and GMP (426). Instead, the flexibility of helix 3 in the yeast GMPK bound to GMP was attributed to it being bound to a single substrate (426). To date, however, no systematic study has been done on the role of helix 3 on the catalytic cycle and conformational changes associated with substrate binding in GMPKs. That said, the location of helix 3, near

the bottom of the GMP-binding domain and adjacent to a hinge region, does make a compelling case for a possible role. **Figure 7.12.** Helix 3 in relation to the GMP—BD and CORE domains. Helix 3 (in purple) is hypothesized to be involved in product release in GMPK. It appears to be positioned as a hinge between the GMP-BD and CORE domains and is proposed to experience an 'unwinding' as GMP-BD moves in response to ligand binding. The red arrow indicates the motion of GMP-BD that 'acts to unwind' helix 3. Order parameters corresponding to this helix are also shown for apo (gray) and GMP-bound (magenta).



In this present work, the complex between GMP and GMPK was investigated. Available data only partially covers the helix but provide some preliminary insights on the dynamics of helix 3 (**Figure 7.12**). A couple of residues within the helix do show reduced flexibility in the ps-ns timescale, although there is also a residue that shows increased flexibility. However, as there is limited data on this region, it is difficult to confirm the proposed role for helix 3 based on its flexibility in the ps-ns timescale upon substrate binding. Interestingly, in the stretch of amino acids from residue 91 to 102, six residues exhibit relaxation dispersion in the bound form, and only three (V99, L100, and D101) are observed in apo (**Figure 7.11**). These three are actually observed in both. As all dispersion profiles were fit into a single cluster, this could mean that this motion corresponds to the global motion that GMPK undergoes. However, more data, particularly at other temperatures, will be necessary to be certain that there are indeed no thermodynamic clusters.

B-factors are typically correlated to the order parameter such that they reflect motion in the ps-ns timescale. Room-temperature X-ray crystallography shows that solution dynamics in this timescale correlate well with order parameters calculated from the generated models (82). However, the relationship between order parameters from solution and B-factors from crystallography is also affected by other factors (83). While a trend typically is observed – that high S^2 is correlated with lower B-factors – factors like the crystal lattice and slower motions in solution may impact the relationship. For instance, elevated B-factors may be explained by multiple conformations that are in slow exchange (83). Thus, an argument could possibly be made that the elevated B-factors observed in yeast GMPK might

correspond to the exchange process observed in this current work. However, this needs to be systematically studied – perhaps by crystallographic studies of GMPK bound to GMP or by looking at solution dynamics of the GMP-bound yeast GMPK. Until then, assigning function to this helix will remain speculation.

What can we do with the AF2 prediction? In the discussion above, it was alluded to that the prediction from AF2 can be used to help design some experiments related to GMPK. Here, two such examples that are related to the current work will be discussed very briefly.

First, the structure of GMPK in the closed form is currently unknown. The available AlphaFold2 prediction can help determine this structure experimentally, whether by providing a model to analyze diffraction data as discussed in Chapter 2, or by using it to analyze sparse NOE restraints from NMR akin to the use of EPR distance restraints in *de novo* structure prediction (*61*, *62*, *439*). Second, the available AF2 structure can be used to design EPR experiments and efficiently (**Figure 7.13**) place spin labels in areas that undergo larger amplitude conformational changes (*87*). EPR-derived ensembles under different substrate conditions can then be obtained and interpreted in tandem with dynamics data from NMR to fully capture the conformational cycle and the dynamics associated with the catalytic cycle of GMPK. Overall, an integrative biology approach will certainly provide a better understanding of GMPK dynamics, especially as a model for a bi-substrate enzyme, and also provide tools that can be used in drug discovery efforts against this potential drug target.

Figure 7.13. Proposed EPR experiments based on AF2 predictions. Residues that undergo large conformational changes based on the apo structure and AF2 prediction can be labelled with spin labels (orange spheres) for EPR/DEER analysis. A few such residues are shown along with the distance changes associated with them. Coupled with NMR data, this will provide further details into the conformational cycle associated with GMPK catalysis.





Spin 1	Domain	Spin 2	Domain	Open (A)	Closed (A)	Change (A)
G46	GMP-BD	N138	LID	31	11	20
T57	GMP-BD	T141	LID	39	18	21
F70	GMP-BD	N138	LID	37	32	5
G46	GMP-BD	T141	LID	34	11	23
T57	GMP-BD	N138	LID	40	23	17
D175	CORE	G46	GMP-BD	34	24	10
M94	CORE	N138	LID	41	34	7

In this chapter, the solution dynamics of human guanylate monophosphate kinase was investigated as well as how these dynamics respond to binding of a single substrate, GMP. Two timescales were probed in this work: the fast picosecond-nanosecond timescale that revealed global ordering upon GMP binding, and the slower microsecond-millisecond timescale that captured a global conformational exchange event in GMPK that is not quenched even with substrate binding. Interestingly, in fact, more residues show dispersion upon addition of GMP. Further experiments involving the ternary complex or binding of ADP or ATP analogs will provide additional information on the role of dynamics and how it is altered throughout the catalytic cycle of GMPK.

CHAPTER 08

NMR: CONCLUDING REMARKS

AND A FEW MORE INTERESTING APPLICATIONS

"A long-term goal of structural biology remains the visualization of molecular structures in their natural context, which is often referred to as incell or in situ structure determination."

Patrick Kramer Nature Structural and Molecular Biology 2020 (440)

Time and time again, the ability of NMR to study biomolecules "near physiological conditions" has arguably been one of its top selling points. The truth, however, is any buffer system will not be a sufficient substitute for the cellular milieu (441). The current methods continue to improve in characterizing protein structure and dynamics *in vitro*, and the long-term goals of *in-cell* structural biology, arguably, are within reach. This work illustrates some applications of NMR and contributes to the already massive collection of studies in the field of *in vitro* structural biology. Thus, the present work exemplifies how deeply engrained NMR is in the field of structural biology, particularly in dynamics and drug discovery, and how heavily these areas rely on, and for years to come will continue to do so, on NMR methods.

This chapter summarizes the present work in as few words as possible, discusses a few more interesting (to the author) NMR-related methods, and offers concluding remarks on the future of NMR and structural biology.

NMR and drug discovery: PRL3 inhibitors. Several inhibitors have been identified for PRL3 but have been deficient in one area or another. The present work adds to this list: limited solubility drugs, candesartan and salirasib, and a weak inhibitor, olsalazine. There is, however, hope in improving these molecules - candesartan derivatives that are more soluble appear to be feasible candidates based on docking results, and olsalazine might prove to be an interesting scaffold. NMR will certainly play a role in characterizing, improving, and screening these and other molecules. Importantly, olsalazine is the very first molecule to be shown to bind to PRL3 – and this was also accomplished by NMR. NMR also served as an excellent validation tool in assessing the binding of other PRL3 inhibitors. Those tested in the present work, however, failed to show binding. While the application of NMR in this work has been limited to binding validation, it is easy to see how it fits in the entire pipeline of drug discovery. Today, it is screening for PRL3 inhibitors, and in the future, it will be characterizing the structure of the complex and assisting in improving the efficacy and specificity of these molecules. Furthermore, in the present work, docking simulations revealed what is possible a novel allosteric site within PRL3 that has not been targeted before. Future studies will validate this site and its 'druggability.' Certainly, NMR will play a role in these studies.

NMR and conformational heterogeneity: MDMX structures. Through the years, MDMX has proven to be a challenging protein to target. Inhibitors identified have not been as successful as the inhibitors of the homologous protein, MDM2. In this work, it has been hypothesized that this difference is due to the variation within the functional conformational subset of the conformational energy landscape and the average structure of MDMX. Data mining on available structures, indeed, reveal that while the different inhibitors identified so far had similar pharmacophores, significant flexibility is observed distal to the binding site, indicating significant conformational heterogeneity of MDMX. Residual dipolar couplings collected on the MDMX:nutlin complex recapitulate the observed differences in residues distal from the p53-binding site. The low Q-factors indicate that the structure of MDMX bound to nutlin is distinct from the available structures despite highly similar pharmacophores. This work, however, is highly preliminary, but shows that the MDMX/MDM2 could serve as an excellent model to study the relationship between dynamics and function. NMR is also very well-suited to tackle this problem, especially if molecular recognition occurring in the hidden time window, which covers about four orders of magnitude of motion from several nanoseconds to tens of microseconds, turns out to be involved in this process.

NMR and dynamics at various timescales: GMPK and substrate binding. GMPK is both an interesting model and therapeutic target. It has eluded drug discovery for so long, but recent data show that GMPK has promise to be an anti-

cancer target. Meanwhile, it is an interesting model for a ternary complex with two distinct substrates. For human GMPK, the solution structure was only recently determined, and remains, to date, the only structure available. There is enough evidence, however, from other organisms and structure predictions, that the conformational changes are conserved - an open-to-closed transition upon substrate binding. However, as evinced by the behavior of helix 3, there are also some differences between the motions of GMPK from various organisms. A systematic study has not been done and this current work provides an initial description of the solution dynamics using NMR relaxation and relaxation dispersion to study picosecond-nanosecond and microsecond-millisecond motion. GMP binding, perhaps as expected, induces ordering in the picosecondnanosecond timescale. Interestingly, CPMG shows more residues exhibiting dispersion in the GMP-bound state. The consequences of these need to be further explored. The current work, to date, is the first application of high-power CPMG to a non-model system and an enzyme. Future work shall characterize the occupancy of the ATP site by an ATP analog or more interestingly, the ternary complex. There is significant work that is expected to come out of this area of study within the next five years, involving state-of-the-art NMR methods of detecting biomolecular motion.

NMR encore: a couple more interesting applications. Clearly, a quick google search, even by the uninitiated, will yield a plethora of advancements and applications in the field of NMR – even within the past few months! During the

COVID pandemic (which is still happening as of this writing), NMR has emerged as a tool to quickly address the current structural and drug discovery needs involving SARS-CoV-2 (442). If anything, this shows the reliability and power of NMR, and also illustrates that even established NMR techniques are still very useful and applicable in today's problems.

Another technique that certainly is not new is ¹⁹F-NMR. The fluorine nucleus can easily be incorporated into proteins, is very sensitive, and has virtually no background (443). Because of these attractive properties, it continues to be used in studies involving protein structure, dynamics, and even drug discovery (444). One of the more interesting applications of ¹⁹F-NMR is in G protein-coupled receptor (GPCR) dynamics. While there are certainly improvements in expression and purification techniques, GPCRs remain very challenging to study. The reemergence of ¹⁹F-NMR, however, has allowed for the study of structural changes and dynamics in response to modulators whether through simple chemical shift perturbation or relaxation dispersion studies (445). This area is clearly one to watch, especially for those interested in GPCR biology. Meanwhile, as mentioned in the beginning of the chapter, *in cell* structural biology remains the goal. ¹⁹F-NMR has also contributed to this aspect and promising approaches should continue to emerge over the next few years (or months!) (446, 447). Possibly, the next application would be ¹⁹F-NMR of GPCR activation inside mammalian cells?

Finally, another area of interest, and perhaps more in line with the current work, is a complete description of protein dynamics in solution. Particularly, it will be interesting to see how computational methodologies and improvements in NMR

will yield a routine protocol to characterize correlated motions and allostery within any protein system (448). Multi-conformer, multi-temperature crystallography has allowed for the visualization of such correlated motions (45, 85). Further improvement in relaxation dispersion studies is possibly how NMR will contribute to this area, along with methods that characterize the full span of timescales of protein motions (99, 117, 139, 145). Of course, this will be coupled with continued improvements in computational methods such as multi-microsecond long MD simulations (76). The era of integrative biology is surely equipped to tackle this outstanding question.

...and beyond. There may be a lot that can be discussed about the future of structural biology, especially in the recent emergence of AI-based tools. However, everything likely boils down to this: The future of structural biology is integrative, dynamic, and *inside the cell*. It is <u>my</u> opinion, that no matter how this future pans out, NMR will be at the center of it.

REFERENCES

1. C. Boesch, Nobel prizes for nuclear magnetic resonance: 2003 and Historical perspectives. *J Magn Reson Imaging*. 20, 177–179 (2004).

2. P. K. Mandal, A. Majumdar, A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts Magn Reso A*. 20A, 1–23 (2004).

3. O. W. Sørensen, G. W. Eich, M. H. Levitt, G. Bodenhausen, R. R. Ernst, Product operator formalism for the description of NMR pulse experiments. *Prog Nucl Mag Res Sp.* 16, 163–192 (1984).

4. J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*. 181, 662–666 (1958).

5. M. F. PERUTZ, M. G. ROSSMANN, A. F. CULLIS, H. MUIRHEAD, G. WILL, A. C. T. NORTH, Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*. 185, 416–422 (1960).

6. A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* 43, W389–W394 (2015).

7. L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server. *Bioinformatics*. 16, 404–405 (2000).

8. T. Smolarczyk, I. Roterman-Konieczna, K. Stapor, Protein Secondary Structure Prediction: A Review of Progress and Directions. *Curr Bioinform*. 15, 90–107 (2019).

9. W. Wardah, M. G. M. Khan, A. Sharma, M. A. Rashid, Protein secondary structure prediction using neural networks and deep learning: a review. *Comput Biol Chem.* 81, 1–8 (2019).

10. C. Fang, Y. Shang, D. Xu, MUFOLD-SS: New deep inception-inside-inception networks for protein secondary structure prediction. *Proteins Struct Funct Bioinform*. 86, 592–598 (2018).

11. V. N. Uversky, A decade and a half of protein intrinsic disorder: Biology still waits for physics. *Protein Sci Publ Protein Soc.* 22, 693–724 (2013).

12. C. J. Oldfield, A. K. Dunker, Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annu Rev Biochem.* 83, 553–584 (2014).

13. V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu Rev Biophys.* 37, 215–246 (2008).

14. P. Tompa, Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci.* 37, 509–516 (2012).

15. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 293, 321–331 (1999).

16. L. Hood, L. Rowen, The Human Genome Project: big science transforms biology and medicine. *Genome Med.* 5, 79–79 (2013).

17. C. A. Orengo, A. E. Todd, J. M. Thornton, From protein structure to function. *Curr Opin Struc Biol.* 9, 374–382 (1999).

18. A. M. J. J. Bonvin, 50 years of PDB: a catalyst in structural biology. *Nat Methods*. 18, 448–449 (2021).

19. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242 (2000).

20. J. Kendrew, R. Parrish, The crystal structure of myoglobin III. sperm-whale myoglobin. *Proc. R. Soc. A.* 238, 305–324 (1957).

21. Y. Shi, A Glimpse of Structural Biology through X-Ray Crystallography. *Cell*. 159, 995–1014 (2014).

22. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biological Crystallogr*. 60, 2126–2132 (2004).

23. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biological Crystallogr*. 66, 213–221 (2010).

24. K. R. Acharya, M. D. Lloyd, The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci.* 26, 10–14 (2005).

25. J. C. Brooks-Bartlett, E. F. Garman, The Nobel Science: One Hundred Years of Crystallography. *Interdiscipl Sci Rev.* 40, 244–264 (2015).

26. J. Birch, D. Axford, J. Foadi, A. Meyer, A. Eckhardt, Y. Thielmann, I. Moraes, The fine art of integral membrane protein crystallisation. *Methods*. 147, 150–162 (2018).

27. T. O. C. Kwan, D. Axford, I. Moraes, Membrane protein crystallography in the era of modern structural biology. *Biochem Soc T.* 48, 2505–2524 (2020).

28. A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, Protein structure determination from NMR chemical shifts. *Proc National Acad Sci.* 104, 9615–9620 (2007).

29. C. Mumenthaler, P. Güntert, W. Braun, K. Wüthrich, Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J Biomol Nmr.* 10, 351–362 (1997).

30. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA. *J Mol Biol.* 319, 209–227 (2002).

31. W. Lee, C. M. Petit, G. Cornilescu, J. L. Stark, J. L. Markley, The AUDANA algorithm for automated protein 3D structure determination from NMR NOE data. *J Biomol Nmr.* 65, 51–57 (2016).

32. A. Bax, S. Grzesiek, Methodological advances in protein NMR. *Accounts Chem Res*. 26, 131–138 (1993).

33. L. E. Kay, NMR studies of protein structure and dynamics. *J Magn Reson*. 173, 193–207 (2005).

34. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc National Acad Sci.* 94, 12366–12371 (1997).

35. J. G. Kempf, J. P. Loria, Protein dynamics from solution NMR. *Cell Biochem Biophys.* 37, 187–211 (2002).

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

37. A. Mittermaier, L. E. Kay, New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science*. 312, 224–228 (2006).

38. A. Sekhar, L. E. Kay, An NMR View of Protein Dynamics in Health and Disease. *Annual Review of Biophysics*. 48, 298–319 (2019).

39. R. Danev, H. Yanagisawa, M. Kikkawa, Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends Biochem Sci.* 44, 837–848 (2019).

40. Y. Cheng, Single-particle cryo-EM—How did it get here and where will it go. *Science*. 361, 876–880 (2018).

41. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. *Nature*. 587, 157–161 (2020).

42. T. Nakane, A. Kotecha, A. Sente, G. McMullan, S. Masiulis, P. M. G. E. Brown, I. T. Grigoras, L. Malinauskaite, T. Malinauskas, J. Miehling, T. Uchański, L. Yu, D. Karia, E. V. Pechnikova, E. de Jong, J. Keizer, M. Bischoff, J. McCormack, P. Tiemeijer, S. W. Hardwick, D. Y. Chirgadze, G. Murshudov, A. R. Aricescu, S. H. W. Scheres, Single-particle cryo-EM at atomic resolution. *Nature*. 587, 152–156 (2020).

43. C. L. Lawson, A. Kryshtafovych, P. D. Adams, P. V. Afonine, M. L. Baker, B. A. Barad, P. Bond, T. Burnley, R. Cao, J. Cheng, G. Chojnowski, K. Cowtan, K. A. Dill, F. DiMaio, D. P. Farrell, J. S. Fraser, M. A. Herzik, S. W. Hoh, J. Hou, L.-W. Hung, M. Igaev, A. P. Joseph, D. Kihara, D. Kumar, S. Mittal, B. Monastyrskyy, M. Olek, C. M. Palmer, A. Patwardhan, A. Perez, J. Pfab, G. D. Pintilie, J. S. Richardson, P. B. Rosenthal, D. Sarkar, L. U. Schäfer, M. F. Schmid, G. F. Schröder, M. Shekhar, D. Si, A. Singharoy, G. Terashi, T. C. Terwilliger, A. Vaiana, L. Wang, Z. Wang, S. A. Wankowicz, C. J. Williams, M. Winn, T. Wu, X. Yu, K. Zhang, H. M. Berman, W. Chiu, Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. *Nat Methods.* 18, 156–164 (2021).

44. L. V. Bock, H. Grubmüller, Effects of cryo-EM cooling on structural ensembles. *Nat Commun.* 13, 1709 (2022).

45. D. A. Keedy, Journey to the center of the protein: allostery from multitemperature multiconformer X-ray crystallography. *Acta Crystallogr Sect D Struct Biology*. 75, 123–137 (2019).

46. R. M. Glaeser, R. J. Hall, Reaching the Information Limit in Cryo-EM of Biological Macromolecules: Experimental Aspects. *Biophys J*. 100, 2331–2337 (2011).

47. M. Wu, G. C. Lander, How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr Opin Struc Biol.* 64, 9–16 (2020).

48. Y. Liu, D. T. Huynh, T. O. Yeates, A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat Commun.* 10, 1864 (2019).

49. A. B. Ward, A. Sali, I. A. Wilson, Integrative Structural Biology. *Science*. 339, 913–915 (2013).

50. G. Masarati, M. Landau, N. Ben-Tal, A. Lupas, M. Kosloff, J. Kosinski, Integrative structural biology in the era of accurate structure prediction. *J Mol Biol*. 433, 167127 (2021).

51. J. Lengyel, E. Hnath, M. Storms, T. Wohlfarth, Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J Struct Funct Genom.* 15, 117–124 (2014).

52. M. P. Rout, A. Sali, Principles for Integrative Structural Biology Studies. *Cell*. 177, 1384–1403 (2019).

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

54. H. van den Bedem, J. S. Fraser, Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat Methods*. 12, 307–318 (2015).

55. K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. *Science*. 338, 1042–1046 (2012).

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

57. J. Pereira, A. J. Simpkin, M. D. Hartmann, D. J. Rigden, R. M. Keegan, A. N. Lupas, High-accuracy protein structure prediction in CASP14. *Proteins Struct Funct Bioinform*. 89, 1687–1699 (2021).

58. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–D444 (2021).

59. D. T. Jones, J. M. Thornton, The impact of AlphaFold2 one year on. *Nat Methods*. 19, 15–20 (2022).

60. A. J. McCoy, M. D. Sammito, R. J. Read, Implications of AlphaFold2 for crystallographic phasing by molecular replacement. *Acta Crystallogr Sect D*. 78, 1–13 (2022).

61. I. Barbarin-Bocahu, M. Graille, The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report. *Acta Crystallogr Sect D*. 78, 517–531 (2022).

62. M. Akdel, D. E. V. Pires, E. P. Pardo, J. Jänes, A. O. Zalevsky, B. Mészáros, P. Bryant, L. L. Good, R. A. Laskowski, G. Pozzati, A. Shenoy, W. Zhu, P. Kundrotas, V. R. Serra, C. H. M. Rodrigues, A. S. Dunham, D. Burke, N. Borkakoti, S. Velankar, A. Frost, K. Lindorff-Larsen, A. Valencia, S. Ovchinnikov, J. Durairaj, D. B. Ascher, J. M. Thornton, N. E. Davey, A. Stein, A. Elofsson, T. I. Croll, P. Beltrao, *Biorxiv*, in press, doi:10.1101/2021.09.26.461876.

63. D. Ringe, G. A. Petsko, Mapping protein dynamics by X-ray diffraction. *Prog Biophysics Mol Biology*. 45, 197–235 (1985).

64. C. B. Anfinsen, E. Haber, Studies on the Reduction and Re-formation of Protein Disulfide Bonds. *J Biol Chem.* 236, 1361–1363 (1961).

65. C. B. Anfinsen, Principles that Govern the Folding of Protein Chains. *Science*. 181, 223–230 (1973).

66. G. Wei, W. Xi, R. Nussinov, B. Ma, Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem Rev.* 116, 6516–6551 (2016).

67. R. Nussinov, P. G. Wolynes, A second molecular biology revolution? The energy landscapes of biomolecular function. *Phys Chem Chem Phys.* 16, 6321–6322 (2014).

68. M. Varadi, S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. Wishart, P. E. Wright, P. Tompa, pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Res.* 42, D326–D335 (2014).

69. H. Frauenfelder, S. Sligar, P. Wolynes, The energy landscape and motions of proteins. *Science*. 254, 1598–1603 (1991).

70. A. Kurakin, Scale-free flow of life: on the biology, economics, and physics of the cell. *Theor Biology Medical Model*. 6, 6–6 (2009).

71. D. Granata, F. Baftizadeh, J. Habchi, C. Galvagnion, A. D. Simone, C. Camilloni, A. Laio, M. Vendruscolo, The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci Rep-uk.* 5, 15449 (2015).

72. A. Cembran, J. Kim, J. Gao, G. Veglia, NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys Chem Chem Phys.* 16, 6508–6518 (2014).

73. P. Wolynes, J. Onuchic, D. Thirumalai, Navigating the Folding Routes. *Science*. 267, 1619–1620 (1995).

74. M. Karplus, J. A. McCammon, Molecular dynamics simulations of biomolecules. *Nat Struct Biol.* 9, 646–652 (2002).

75. R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, D. E. Shaw, Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu Rev Biophys.* 41, 429–452 (2012).

76. K. S. Shim, B. Greskamp, B. Towles, B. Edwards, J. P. Grossman, D. E. Shaw, The Specialized High-Performance Network on Anton 3. *Arxiv* (2022).

77. A. Palmer, NMR probes of molecular dynamics: overview and comparison with other techniques. *Annual Review of Biophysics and Biomolecular Structure*. 30, 129–155 (2001).

78. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*. 330, 341–346 (2010).

79. E. S. O'Brien, A. J. Wand, K. A. Sharp, On the ability of molecular dynamics force fields to recapitulate NMR derived protein side chain order parameters. *Protein Sci.* 25, 1156–1160 (2016).

80. M. Tiberti, E. Papaleo, T. Bengtsen, W. Boomsma, K. Lindorff-Larsen, ENCORE: Software for Quantitative Ensemble Comparison. *Plos Comput Biol.* 11, e1004415 (2015).

81. T. M. Sabo, J. O. Trent, D. Lee, Population shuffling between ground and high energy excited states. *Protein Sci.* 24, 1714–1719 (2015).

82. R. B. Fenwick, H. van den Bedem, J. S. Fraser, P. E. Wright, Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc National Acad Sci.* 111, E445–E454 (2014).

83. R. Powers, G. M. Clore, D. S. Garrett, A. M. Gronenborn, Relationships Between the Precision of High-Resolution Protein NMR Structures, Solution-Order Parameters, and Crystallographic B Factors. *J Magnetic Reson Ser B*. 101, 325–327 (1993).

84. B. T. Riley, S. A. Wankowicz, S. H. P. Oliveira, G. C. P. Zundert, D. W. Hogan, J. S. Fraser, D. A. Keedy, H. Bedem, qFit 3: Protein and ligand multiconformer modeling for

X-ray crystallographic and single-particle cryo-EM density maps. *Protein Sci.* 30, 270–285 (2021).

85. D. A. Keedy, L. R. Kenner, M. Warkentin, R. A. Woldeyes, J. B. Hopkins, M. C. Thompson, A. S. Brewster, A. H. V. Benschoten, E. L. Baxter, M. Uervirojnangkoorn, S. E. McPhillips, J. Song, R. Alonso-Mori, J. M. Holton, W. I. Weis, A. T. Brunger, S. M. Soltis, H. Lemke, A. Gonzalez, N. K. Sauter, A. E. Cohen, H. van den Bedem, R. E. Thorne, J. S. Fraser, Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *Elife.* 4, e07574 (2015).

86. W. L. Hubbell, C. J. López, C. Altenbach, Z. Yang, Technological advances in sitedirected spin labeling of proteins. *Curr Opin Struc Biol.* 23, 725–733 (2013).

87. G. Jeschke, The contribution of modern EPR to structural biology. *Emerg Top Life Sci.* 2, 9–18 (2018).

88. H. S. Mchaourab, P. R. Steed, K. Kazmier, Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy. *Structure*. 19, 1549–1561 (2011).

89. H. Sanabria, D. Rodnin, K. Hemmen, T.-O. Peulen, S. Felekyan, M. R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke, C. A. M. Seidel, Resolving dynamics and function of transient states in single enzyme molecules. *Nat Commun.* 11, 1231 (2020).

90. B. Roux, S. M. Islam, Restrained-Ensemble Molecular Dynamics Simulations Based on Distance Histograms from Double Electron–Electron Resonance Spectroscopy. *J Phys Chem B*. 117, 4733–4739 (2013).

91. F. Torricella, A. Pierro, E. Mileo, V. Belle, A. Bonucci, Nitroxide spin labels and EPR spectroscopy: A powerful association for protein dynamics studies. *Biochimica Et Biophysica Acta Bba - Proteins Proteom.* 1869, 140653 (2021).

92. K. Okamoto, Y. Sako, Recent advances in FRET for the study of protein interactions and dynamics. *Current Opinion in Structural Biology*. 46, 16–23 (2017).

93. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat Methods*. 5, 507–516 (2008).

94. H. Mazal, G. Haran, Single-molecule FRET methods to study the dynamics of proteins at work. *Curr Opin Biomed Eng.* 12, 8–17 (2019).

95. M. F. Peter, C. Gebhardt, R. Mächtel, J. Glaenzer, G. H. Thomas, T. Cordes, G. Hagelueken, *Biorxiv*, in press, doi:10.1101/2020.11.23.394080.

96. M. Faini, F. Stengel, R. Aebersold, The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J Am Soc Mass Spectr.* 27, 966–974 (2016).

97. L. E. Kay, Faseb J, in press, doi:10.1096/fasebj.2018.32.1_supplement.475.2.

98. M. Osawa, K. Takeuchi, T. Ueda, N. Nishida, I. Shimada, Functional dynamics of proteins revealed by solution NMR. *Curr Opin Struc Biol.* 22, 660–669 (2012).

99. T. Ikeya, D. Ban, D. Lee, Y. Ito, K. Kato, C. Griesinger, Solution NMR views of dynamical ordering of biomacromolecules. *Biochimica Et Biophysica Acta Bba - Gen Subj.* 1862, 287–306 (2018).

100. G. M. Clore, A. M. Gronenborn, New methods of structure refinement for macromolecular structure determination by NMR. *Proceedings of the National Academy of Science*. 95, 5891–5898 (1998).

101. P. Güntert, Automated structure determination from NMR spectra. *Eur Biophys J.* 38, 129 (2008).

102. E. M. Yilmaz, P. Güntert, NMR structure calculation for all small molecule ligands and non-standard residues from the PDB Chemical Component Dictionary. *J Biomol Nmr*. 63, 21–37 (2015).

103. A. K. Downing, P. Güntert, Protein NMR Techniques. *Methods Mol Biology Clifton* NJ. 278, 353–378 (2004).

104. V. Kanelis, J. D. Forman-Kay, L. E. Kay, Multidimensional NMR Methods for Protein Structure Determination. *Iubmb Life*. 52, 291–302 (2001).

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

106. M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp.* 34, 93–158 (1999).

107. M. Nilges, Calculation of Protein Structures with Ambiguous Distance Restraints. Automated Assignment of Ambiguous NOE Crosspeaks and Disulphide Connectivities. *J Mol Biol.* 245, 645–660 (1995).

108. I. Solomon, Relaxation Processes in a System of Two Spins. *Phys Rev.* 99, 559–565 (1955).

109. D. Neuhaus, Nuclear Overhauser effect. *Encyclopedia of Magnetic Resonance*, 1–12 (2007).
110. R. S. Lipsitz, N. Tjandra, RESIDUAL DIPOLAR COUPLINGS IN NMR STRUCTURE ANALYSIS*1. *Annu Rev Bioph Biom.* 33, 387–413 (2004).

111. J. H. Prestegard, C. M. Bougault, A. I. Kishore, Residual Dipolar Couplings in Structure Determination of Biomolecules. *Chem Rev.* 104, 3519–3540 (2004).

112. C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, The Xplor-NIH NMR molecular structure determination package. *J Magn Reson*. 160, 65–73 (2003).

113. B. Schuler, H. Hofmann, Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr Opin Struc Biol.* 23, 36–47 (2013).

114. D. Khago, I. J. Fucci, R. A. Byrd, The Role of Conformational Dynamics in the Recognition and Regulation of Ubiquitination. *Molecules*. 25, 5933 (2020).

115. F.-A. Chao, R. A. Byrd, Protein dynamics revealed by NMR relaxation methods. *Emerg Top Life Sci.* 2, 93–105 (2018).

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

117. D. Ban, C. A. Smith, B. L. de Groot, C. Griesinger, D. Lee, Recent advances in measuring the kinetics of biomolecules by NMR relaxation dispersion spectroscopy. *Arch Biochem Biophys.* 628, 81–91 (2017).

118. R. Ghose, K. Huang, J. H. Prestegard, Measurement of Cross Correlation between Dipolar Coupling and Chemical Shift Anisotropy in the Spin Relaxation of 13C, 15N-Labeled Proteins. *J Magn Reson*. 135, 487–499 (1998).

119. J. J. Chou, D. A. Case, A. Bax, Insights into the Mobility of Methyl-Bearing Side Chains in Proteins from 3 J CC and 3 J CN Couplings. *J Am Chem Soc.* 125, 8959–8966 (2003).

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

121. A. Sekhar, L. E. Kay, NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc National Acad Sci.* 110, 12867–12874 (2013).

122. D. Ban, T. Sabo, C. Griesinger, D. Lee, Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra-tc Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy. *Molecules*. 18, 11904–11937 (2013). 123. N.-A. Lakomek, O. F. Lange, K. F. A. Walter, C. Farès, D. Egger, P. Lunkenheimer, J. Meiler, H. Grubmüller, S. Becker, B. L. de Groot, C. Griesinger, Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochem Soc T*. 36, 1433–1437 (2008).

124. B. Vögeli, L. Yao, Correlated Dynamics between Protein HN and HC Bonds Observed by NMR Cross Relaxation. *J Am Chem Soc.* 131, 3668–3678 (2009).

125. B. Vögeli, L. Yao, A. Bax, Protein backbone motions viewed by intraresidue and sequential HN–Hα residual dipolar couplings. *J Biomol Nmr*. 41, 17 (2008).

126. G. M. Clore, J. Iwahara, Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem Rev.* 109, 4108–4139 (2009).

127. P. Robustelli, K. A. Stafford, A. G. Palmer, Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J Am Chem Soc.* 134, 6365–6374 (2012).

128. R. B. Fenwick, S. Esteban-Martín, B. Richter, D. Lee, K. F. A. Walter, D. Milovanovic, S. Becker, N. A. Lakomek, C. Griesinger, X. Salvatella, Weak Long-Range Correlated Motions in a Surface Patch of Ubiquitin Involved in Molecular Recognition. *J Am Chem Soc.* 133, 10336–10339 (2011).

129. J. R. Tolman, K. Ruan, NMR Residual Dipolar Couplings as Probes of Biomolecular Dynamics. *Chem Rev.* 106, 1720–1736 (2006).

130. J. Meiler, J. J. Prompers, W. Peti, C. Griesinger, R. Brüschweiler, Model-Free Approach to the Dynamic Interpretation of Residual Dipolar Couplings in Globular Proteins. *J Am Chem Soc.* 123, 6098–6107 (2001).

131. T. M. Sabo, C. A. Smith, D. Ban, A. Mazur, D. Lee, C. Griesinger, ORIUM: Optimized RDC-based Iterative and Unified Model-free analysis. *J Biomol Nmr*. 58, 287–301 (2014).

132. R. Konrat, NMR contributions to structural dynamics studies of intrinsically disordered proteins ☆. *J Magn Reson*. 241, 74–85 (2014).

133. Q. Gong, R. Ishima, 15N–{1H} NOE experiment at high magnetic field strengths. *J Biomol Nmr*. 37, 147–157 (2007).

134. F. Ferrage, D. Cowburn, R. Ghose, Accurate Sampling of High-Frequency Motions in Proteins by Steady-State 15N–{1H} Nuclear Overhauser Effect Measurements in the Presence of Cross-Correlated Relaxation. *J Am Chem Soc.* 131, 6048–6049 (2009).

135. E. J. d'Auvergne, P. R. Gooley, The use of model selection in the model-free analysis of protein dynamics. *J Biomol Nmr*. 25, 25–39 (2003).

136. P. Dosset, J.-C. Hus, M. Blackledge, D. Marion, Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol Nmr*. 16, 23–28 (2000).

137. T. Crawley, A. G. P. III, Bootstrap aggregation for model selection in the model-free formalism. *Magnetic Reson.* 2, 251–264 (2021).

138. Ł. Jaremko, M. Jaremko, M. Nowakowski, A. Ejchart, The Quest for Simplicity: Remarks on the Free-Approach Models. *J Phys Chem B*. 119, 11978–11987 (2015).

139. O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B. L. de Groot, Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science*. 320, 1471–1475 (2008).

140. S. Forsén, R. A. Hoffman, Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance. *J Chem Phys.* 39, 2892– 2901 (1963).

141. P. Vallurupalli, G. Bouvignies, L. E. Kay, Studying "Invisible" Excited Protein States in Slow Exchange with a Major State Conformation. *J Am Chem Soc.* 134, 8148–8161 (2012).

142. Y. Gu, A. L. Hansen, Y. Peng, R. Brüschweiler, Rapid Determination of Fast Protein Dynamics from NMR Chemical Exchange Saturation Transfer Data. *Angew Chem-ger Edit.* 128, 3169–3171 (2016).

143. C. Wang, M. J. Grey, A. G. Palmer, CPMG sequences with enhanced sensitivity to chemical exchange. *J Biomol Nmr*. 21, 361–366 (2001).

144. R. Ishima, D. A. Torchia, Estimating the time scale of chemical exchange of proteins from measurements of transverse relaxation rates in solution. *J Biomol Nmr*. 14, 369–372 (1999).

145. D. Ban, A. D. Gossert, K. Giller, S. Becker, C. Griesinger, D. Lee, Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J Magn Reson*. 221, 1–4 (2012).

146. A. Mazur, B. Hammesfahr, C. Griesinger, D. Lee, M. Kollmar, ShereKhan calculating exchange parameters in relaxation dispersion data from CPMG experiments. *Bioinformatics*. 29, 1819–1820 (2013).

147. Z. Luz, S. Meiboom, Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *J Chem Phys.* 39, 366–370 (1963).

148. P. J. Farber, A. Mittermaier, Relaxation dispersion NMR spectroscopy for the study of protein allostery. *Biophysical Rev.* 7, 191–200 (2015).

149. J. P. Carver, R. E. Richards, A general two-site solution for the chemical exchange produced dependence of T2 upon the carr-Purcell pulse separation. *J Magnetic Reson 1969.* 6, 89–105 (1972).

150. H. M. McConnell, Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys.* 28, 430–431 (1958).

151. X. Xiang, A. L. Hansen, L. Yu, G. Jameson, L. Bruschweiler-Li, C. Yuan, R. Brüschweiler, Observation of Sub-Microsecond Protein Methyl-Side Chain Dynamics by Nanoparticle-Assisted NMR Spin Relaxation. *J Am Chem Soc.* 143, 13593–13604 (2021).

152. S. Wardenfelt, X. Xiang, M. Xie, L. Yu, L. Bruschweiler-Li, R. Brüschweiler, Broadband Dynamics of Ubiquitin by Anionic and Cationic Nanoparticle Assisted NMR Spin Relaxation. *Angewandte Chemie Int Ed.* 60, 148–152 (2021).

153. K. Ruan, J. R. Tolman, Composite Alignment Media for the Measurement of Independent Sets of NMR Residual Dipolar Couplings. *J Am Chem Soc.* 127, 15032–15033 (2005).

154. L. Yao, J. Ying, A. Bax, Improved accuracy of 15N–1H scalar and residual dipolar couplings from gradient-enhanced IPAP-HSQC experiments on protonated proteins. *J Biomol Nmr.* 43, 161–170 (2009).

155. G. Bodenhausen, D. J. Ruben, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett.* 69, 185–189 (1980).

156. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Mag Res Sp.* 73, 1–16 (2013).

157. C. A. Waudby, A. Ramos, L. D. Cabrita, J. Christodoulou, Two-Dimensional NMR Lineshape Analysis. *Sci Rep-uk.* 6, 24826 (2016).

158. J. Xu, S. R. V. Doren, Tracking Equilibrium and Nonequilibrium Shifts in Data with TREND. *Biophys J.* 112, 224–233 (2017).

159. R. Horst, K. A. Farley, B. L. Kormos, J. M. Withka, NMR spectroscopy: the swiss army knife of drug discovery. *J Biomol Nmr*. 74, 509–519 (2020).

160. M. Pellecchia, D. S. Sem, K. Wüthrich, Nmr in drug discovery. *Nat Rev Drug Discov.* 1, 211–219 (2002).

161. T. Diercks, M. Coles, H. Kessler, Applications of NMR in drug discovery. *Curr Opin Chem Biol.* 5, 285–291 (2001).

162. A.-H. Emwas, K. Szczepski, B. G. Poulson, K. Chandra, R. T. McKay, M. Dhahri, F. Alahmari, L. Jaremko, J. I. Lachowicz, M. Jaremko, NMR as a "Gold Standard" Method in Drug Design and Discovery. *Molecules*. 25, 4597 (2020).

163. D. A. Erlanson, R. S. McDowell, T. O'Brien, Fragment-Based Drug Discovery. J Med Chem. 47, 3463–3482 (2004).

164. M. J. Harner, A. O. Frank, S. W. Fesik, Fragment-based drug discovery using NMR spectroscopy. *J Biomol Nmr*. 56, 65–75 (2013).

165. C. W. Murray, D. C. Rees, The rise of fragment-based drug discovery. *Nat Chem.* 1, 187–192 (2009).

166. S. Shuker, P. Hajduk, R. Meadows, S. Fesik, Discovering high affinity ligands for proteins: SAR by NMR. *Science*. 274, 1531–1534 (1996).

167. J. Klages, M. Coles, H. Kessler, NMR-based screening: a powerful tool in fragment-based drug discovery. *Analyst.* 132, 692–705 (2006).

168. A. Ross, H. Senn, Automation of Biomolecular NMR Screening. *Curr Top Med Chem.* 3, 55–67 (2003).

169. P. Hajduk, D. Burns, Integration of NMR and high-throughput screening. *Combinatorial Chemistry & High Throughput Screening*. 5, 613–621 (2002).

170. A. Viegas, J. Manso, F. L. Nobrega, E. J. Cabrita, Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J Chem Educ.* 88, 990–994 (2011).

171. J. Angulo, P. M. Nieto, STD-NMR: application to transient interactions between biomolecules—a quantitative approach. *Eur Biophys J*. 40, 1357–1369 (2011).

172. R. Huang, A. Bonnichon, T. D. W. Claridge, I. K. H. Leung, Protein-ligand binding affinity determination by the waterLOGSY method: An optimised approach considering ligand rebinding. *Sci Rep-uk.* 7, 43727 (2017).

173. C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J Biomol Nmr.* 21, 349–359 (2001).

174. G. L. Rosano, E. S. Morales, E. A. Ceccarelli, New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Sci.* 28, 1412–1422 (2019).

175. W. H. Brondyk, Selecting an Appropriate Method for Expressing a Recombinant Protein. *Methods Enzymol.* 463, 131–147 (2009).

176. R. Chen, Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnol Adv.* 30, 1102–1107 (2012).

177. G. L. Rosano, E. A. Ceccarelli, Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol.* 5, 172 (2014).

178. S. B. Azatian, N. Kaur, M. P. Latham, Increasing the buffering capacity of minimal media leads to higher protein yield. *J Biomol Nmr*. 73, 11–17 (2019).

179. R. J. Noel, W. S. Reznikoff, Structural Studies of lacUV5-RNA Polymerase Interactions in Vitro ETHYLATION INTERFERENCE AND MISSING NUCLEOSIDE ANALYSIS*. *J Biol Chem.* 275, 7708–7712 (2000).

180. M. Cai, Y. Huang, J. Lloyd, R. Craigie, G. M. Clore, A simple and cost-effective protocol for high-yield expression of deuterated and selectively isoleucine/leucine/valine methyl protonated proteins in Escherichia coli grown in shaker flasks. *J Biomol Nmr*. 75, 83–87 (2021).

181. G. Montelione, B. Lyons, D. Emerson, M. Tashiro, An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence specific resonance assignments of isotopically. *Journal of the American Chemical Society*. 114, 10974–10975 (1992).

182. S. Grzesiek, J. Anglister, A. Bax, Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. *J Magnetic Reson Ser B*. 101, 114–119 (1993).

183. N. Khan, P. P. Shah, D. Ban, P. Trigo-Mouriño, M. G. Carneiro, L. DeLeeuw, W. L. Dean, J. O. Trent, L. J. Beverly, M. Konrad, D. Lee, T. M. Sabo, Solution structure and functional investigation of human guanylate kinase reveals allosteric networking and a crucial role for the enzyme in cancer. *J Biol Chem.* 294, 11920–11933 (2019).

184. M. Bastos, A. Velazquez-Campoy, Isothermal titration calorimetry (ITC): a standard operating procedure (SOP). *Eur Biophys J*. 50, 363–371 (2021).

185. H. Zhao, G. Piszczek, P. Schuck, SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods*. 76, 137–148 (2015).

186. J. Krishnamoorthy, S. Mohanty, Open-ITC: an alternate computational approach to analyze the isothermal titration calorimetry data of complex binding mechanisms. *J Mol Recognit.* 24, 1056–1066 (2011).

187. H. Duvvuri, L. C. Wheeler, M. J. Harms, pytc: Open-Source Python Software for Global Analyses of Isothermal Titration Calorimetry Data. *Biochemistry-us.* 57, 2578–2583 (2018).

188. M. Jerabek-Willemsen, T. André, R. Wanner, H. M. Roth, S. Duhr, P. Baaske, D. Breitsprecher, MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct.* 1077, 101–113 (2014).

189. J. M. McDonnell, Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Curr Opin Chem Biol.* 5, 572–577 (2001).

190. A. Sultana, J. E. Lee, *Curr Protoc Protein Sci*, in press, doi:10.1002/0471140864.ps1925s79.

191. J. Tellinghuisen, Optimizing isothermal titration calorimetry protocols for the study of 1:1 binding: Keeping it simple. *Biochimica Et Biophysica Acta Bba - Gen Subj.* 1860, 861–867 (2016).

192. J. Tellinghuisen, Optimizing Experimental Parameters in Isothermal Titration Calorimetry. *J Phys Chem B*. 109, 20027–20035 (2005).

193. J. Broecker, C. Vargas, S. Keller, Revisiting the optimal c value for isothermal titration calorimetry. *Anal Biochem.* 418, 307–309 (2011).

194. P. Cohen, D. R. Alessi, Kinase Drug Discovery – What's Next in the Field? *Acs Chem Biol.* 8, 96–104 (2013).

195. M. K. Tarrant, P. A. Cole, The Chemical Biology of Protein Phosphorylation. *Annu Rev Biochem.* 78, 797–825 (2009).

196. L. N. Johnson, The regulation of protein phosphorylation. *Biochem Soc T*. 37, 627–641 (2009).

197. F. Ardito, M. Giuliani, D. Perrone, G. Troiano, L. L. Muzio, The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int J Mol Med.* 40, 271–280 (2017).

198. V. Singh, M. Ram, R. Kumar, R. Prasad, B. K. Roy, K. K. Singh, Phosphorylation: Implications in Cancer. *Protein J.* 36, 1–6 (2017).

199. S. J. Humphrey, D. E. James, M. Mann, Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol Metabolism*. 26, 676–687 (2015).

200. T. Hunter, Signaling—2000 and Beyond. Cell. 100, 113–127 (2000).

201. L. R. Bollu, A. Mazumdar, M. I. Savage, P. H. Brown, Molecular Pathways: Targeting Protein Tyrosine Phosphatases in Cancer. *Clin Cancer Res.* 23, 2136–2142 (2017).

202. Z.-Y. Zhang, Drugging the Undruggable: Therapeutic Potential of Targeting Protein Tyrosine Phosphatases. *Accounts Chem Res.* 50, 122–129 (2016).

203. Q. Yang, J. C. Langston, Y. Tang, M. F. Kiani, L. E. Kilpatrick, The Role of Tyrosine Phosphorylation of Protein Kinase C Delta in Infection and Inflammation. *Int J Mol Sci.* 20, 1498 (2019).

204. P. Viatour, M.-P. Merville, V. Bours, A. Chariot, Phosphorylation of NF-κB and IκB proteins: implications in cancer and inflammation. *Trends Biochem Sci.* 30, 43–52 (2005).

205. D. Hanahan, R. A. Weinberg, The Hallmarks of Cancer. Cell. 100, 57-70 (2000).

206. P. J. Roberts, C. J. Der, Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*. 26, 3291–3310 (2007).

207. E. Appella, C. Anderson, Post-translational modifications and activation of p53 by genotoxic stresses. *European Journal of Biochemistry*. 268, 2764–2772 (2001).

208. O. M. Fischer, S. Streit, S. Hart, A. Ullrich, Beyond Herceptin and Gleevec. *Curr Opin Chem Biol.* 7, 490–495 (2003).

209. P. Cohen, Protein kinases — the major drug targets of the twenty-first century? *Nat Rev Drug Discov.* 1, 309–315 (2002).

210. G. Giamas, J. Stebbing, C. E. Vorgias, U. Knippschild, Protein kinases as targets for cancer treatment. *Pharmacogenomics*. 8, 1005–1016 (2007).

211. R. Roskoski, Properties of FDA-approved small molecule protein kinase inhibitors: a 2021 update. *Pharmacol Res.* 165, 105463 (2021).

212. R. Roskoski, A historical overview of protein kinases and their targeted small molecule inhibitors. *Pharmacol Res.* 100, 1–23 (2015).

213. P. Cohen, D. Cross, P. A. Jänne, Kinase drug discovery 20 years after imatinib: progress and future directions. *Nat Rev Drug Discov*. 20, 1–19 (2021).

214. J. S. Lazo, E. R. Sharlow, Drugging Undruggable Molecular Cancer Targets. *Annu Rev Pharmacol.* 56, 1–18 (2015).

215. J. S. Lazo, K. E. McQueeney, E. R. Sharlow, New Approaches to Difficult Drug Targets: The Phosphatase Story. *Slas Discov.* 22, 1071–1083 (2017).

216. A. Thomson, C. Bonham, A. Zeevi, Mode of action of tacrolimus (FK506): molecular and cellular mechanisms. *Therapeutic Drug Monitoring*. 17, 584–591 (1995).

217. V. Reiterer, K. Pawłowski, G. Desrochers, A. Pause, H. J. Sharpe, H. Farhan, The dead phosphatases society: a review of the emerging roles of pseudophosphatases. *Febs J*. 287, 4198–4220 (2020).

218. N. K. Tonks, Protein tyrosine phosphatases – from housekeeping enzymes to master regulators of signal transduction. *Febs J.* 280, 346–378 (2013).

219. S. Fahs, P. Lujan, M. Köhn, Approaches to Study Phosphatases. *Acs Chem Biol*. 11, 2944–2961 (2016).

220. S. D. Munter, M. Köhn, M. Bollen, Challenges and Opportunities in the Development of Protein Phosphatase-Directed Therapeutics. *Acs Chem Biol.* 8, 36–45 (2013).

221. I. Nasa, A. N. Kettenbach, Coordination of Protein Kinase and Phosphoprotein Phosphatase Activities in Mitosis. *Frontiers Cell Dev Biology*. 6, 30 (2018).

222. A. J. Barr, Protein tyrosine phosphatases as drug targets: strategies and challenges of inhibitor development. *Future Med Chem.* 2, 1563–1576 (2010).

223. J. S. Lazo, K. E. McQueeney, E. R. Sharlow, New Approaches to Difficult Drug Targets: The Phosphatase Story. *SLAS Discovery*. 22, 1071–1083 (2017).

224. S. Müller, A. Chaikuad, N. S. Gray, S. Knapp, The ins and outs of selective kinase inhibitor development. *Nat Chem Biol.* 11, 818–821 (2015).

225. M. A. Bogoyevitch, D. P. Fairlie, A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discov Today*. 12, 622–633 (2007).

226. K. Umezawa, I. Kii, Druggable Transient Pockets in Protein Kinases. *Molecules*. 26, 651 (2021).

227. J. J.-L. Liao, Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *J Med Chem.* 50, 409–424 (2007).

228. M. Köhn, Turn and Face the Strange: A New View on Phosphatases. *Acs Central Sci.* 6, 467–477 (2020).

229. S. M. Stanford, N. Bottini, Targeting Tyrosine Phosphatases: Time to End the Stigma. *Trends Pharmacol Sci.* 38, 524–540 (2017).

230. C. V. Dang, E. P. Reddy, K. M. Shokat, L. Soucek, Drugging the "undruggable" cancer targets. *Nat Rev Cancer*. 17, 502–508 (2017).

231. Z.-Y. Zhang, Protein tyrosine phosphatases: prospects for therapeutics. *Curr Opin Chem Biol.* 5, 416–423 (2001).

232. R. H. van Huijsduijnen, A. Bombrun, D. Swinnen, Selecting protein tyrosine phosphatases as drug targets. *Drug Discov Today*. 7, 1013–1019 (2002).

233. J. S. Lazo, K. E. McQueeney, J. C. Burnett, P. Wipf, E. R. Sharlow, Small molecule targeting of PTPs in cancer. *Int J Biochem Cell Biology*. 96, 171–181 (2018).

234. V. V. Vintonyak, H. Waldmann, D. Rauh, Using small molecules to target protein phosphatases. *Bioorgan Med Chem.* 19, 2145–2155 (2011).

235. J. Vainonen, M. Momeny, J. Westermarck, Druggable cancer phosphatases. *Science Translational Medicine*. 13, 1–13 (2021).

236. D. Barford, A. J. Flint, N. K. Tonks, Crystal structure of human protein tyrosine phosphatase 1B. *Science*. 263, 1397–1404 (1994).

237. M. Feldhammer, N. Uetani, D. Miranda-Saavedra, M. L. Tremblay, PTP1B: A simple enzyme for a complex world. *Crit Rev Biochem Mol.* 48, 430–445 (2013).

238. N. K. Tonks, PTP1B: From the sidelines to the front lines! *Febs Lett.* 546, 140–148 (2003).

239. J. Bakke, F. G. Haj, Protein-tyrosine phosphatase 1B substrates and metabolic regulation. *Semin Cell Dev Biol.* 37, 58–65 (2015).

240. N. Krishnan, C. A. Bonham, I. A. Rus, O. K. Shrestha, C. M. Gauss, A. Haque, A. Tocilj, L. Joshua-Tor, N. K. Tonks, Harnessing insulin- and leptin-induced oxidation of PTP1B for therapeutic development. *Nat Commun.* 9, 283 (2018).

241. L. Klaman, O. Boss, O. Peroni, J. Kim, J. Martino, J. Zbolotny, N. Moghal, M. Lubkin, Y.-B. Kim, A. Sharpe, A. Stricker-Krongrad, G. Shulman, B. Neel, B. Kahn, Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase-1B-deficient mice. *Molecular and Cellular Biology*. 20, 5479–5489 (2000).

242. A. Cheng, N. Uetani, P. D. Simoncic, V. P. Chaubey, A. Lee-Loy, C. J. McGlade, B. P. Kennedy, M. L. Tremblay, Attenuation of Leptin Action and Regulation of Obesity by Protein Tyrosine Phosphatase 1B. *Dev Cell*. 2, 497–503 (2002).

243. M. Elchebly, P. Payette, E. Michaliszyn, W. Cromlish, S. Collins, A. L. Loy, D. Normandin, A. Cheng, J. Himms-Hagen, C.-C. Chan, C. Ramachandran, M. Gresser, M.

Tremblay, B. Kennedy, Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science*. 283, 1544–1548 (1999).

244. N. K. Tonks, S. K. Muthuswamy, A Brake Becomes an Accelerator: PTP1B—A New Therapeutic Target for Breast Cancer. *Cancer Cell*. 11, 214–216 (2007).

245. M. Stuible, K. M. Doody, M. L. Tremblay, PTP1B and TC-PTP: regulators of transformation and tumorigenesis. *Cancer Metast Rev.* 27, 215–230 (2008).

246. L. Lessard, M. Stuible, M. L. Tremblay, The two faces of PTP1B in cancer. *Biochimica Et Biophysica Acta Bba - Proteins Proteom.* 1804, 613–619 (2010).

247. M. Tartaglia, C. M. Niemeyer, A. Fragale, X. Song, J. Buechner, A. Jung, K. Hählen, H. Hasle, J. D. Licht, B. D. Gelb, Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet.* 34, 148–150 (2003).

248. R. J. Chan, G.-S. Feng, PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood*. 109, 862–867 (2006).

249. M. Bentires-Alj, J. G. Paez, F. David, H. Keilhack, B. Halmos, K. Naoki, J. Maris, A. Richardson, A. Bardelli, D. Sugarbaker, W. Richards, J. Du, L. Girard, J. Minna, M. Loh, D. Fisher, V. Velculescu, B. Vogelstein, M. Meyerson, W. Sellers, B. Neel, Activating mutations in the Noonan syndrome-associated SHP2/PTPN11 gene in human solid tumors and adult acute myelogenous leukemia. *Cancer Research*. 64, 8816–8820 (2004).

250. Y.-N. P. Chen, M. J. LaMarche, H. M. Chan, P. Fekkes, J. Garcia-Fortanet, M. G. Acker, B. Antonakos, C. H.-T. Chen, Z. Chen, V. G. Cooke, J. R. Dobson, Z. Deng, F. Fei, B. Firestone, M. Fodor, C. Fridrich, H. Gao, D. Grunenfelder, H.-X. Hao, J. Jacob, S. Ho, K. Hsiao, Z. B. Kang, R. Karki, M. Kato, J. Larrow, L. R. L. Bonte, F. Lenoir, G. Liu, S. Liu, D. Majumdar, M. J. Meyer, M. Palermo, L. Perez, M. Pu, E. Price, C. Quinn, S. Shakya, M. D. Shultz, J. Slisz, K. Venkatesan, P. Wang, M. Warmuth, S. Williams, G. Yang, J. Yuan, J.-H. Zhang, P. Zhu, T. Ramsey, N. J. Keen, W. R. Sellers, T. Stams, P. D. Fortin, Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature*. 535, 148–152 (2016).

251. Z.-X. Jiang, Z.-Y. Zhang, Targeting PTPs with small molecule inhibitors in cancer treatment. *Cancer Metast Rev.* 27, 263–272 (2008).

252. K. L. Mohn, T. M. Laz, J. C. Hsu, A. E. Melby, R. Bravo, R. Taub, The immediateearly growth response in regenerating liver and insulin-stimulated H-35 cells: comparison with serum-stimulated 3T3 cells and identification of 41 novel immediate-early genes. *Mol Cell Biol.* 11, 381–390 (1991). 253. S. Saha, A. Bardelli, P. Buckhaults, V. E. Velculescu, C. Rago, B. St. Croix, K. E. Romans, M. A. Choti, C. Lengauer, K. W. Kinzler, B. Vogelstein, A Phosphatase Associated with Metastasis of Colorectal Cancer. *Science*. 294, 1343–1346 (2001).

254. R. Beekman, M. Valkhof, S. J. Erkeland, E. Taskesen, V. Rockova, J. K. Peeters, P. J. M. Valk, B. Löwenberg, I. P. Touw, Retroviral Integration Mutagenesis in Mice and Comparative Analysis in Human AML Identify Reduced PTP4A3 Expression as a Prognostic Indicator. *Plos One*. 6, e26537 (2011).

255. N. Dai, A.-P. Lu, C.-C. Shou, J.-Y. Li, Expression of phosphatase regenerating liver 3 is an independent prognostic indicator for gastric cancer. *World J Gastroentero*. 15, 1499–1505 (2009).

256. D. C. Bessette, D. Qiu, C. J. Pallen, PRL PTPs: mediators and markers of cancer progression. *Cancer Metast Rev.* 27, 231–252 (2008).

257. A. Mayinuer, M. Yasen, K. Mogushi, G. Obulhasim, M. Xieraili, A. Aihara, S. Tanaka, H. Mizushima, H. Tanaka, S. Arii, Upregulation of Protein Tyrosine Phosphatase Type IVA Member 3 (PTP4A3/PRL-3) is Associated with Tumor Differentiation and a Poor Prognosis in Human Hepatocellular Carcinoma. *Ann Surg Oncol.* 20, 305–317 (2013).

258. S. Qu, B. Liu, X. Guo, H. Shi, M. Zhou, L. Li, S. Yang, X. Tong, H. Wang, Independent oncogenic and therapeutic significance of phosphatase PRL-3 in FLT3-ITD– negative acute myeloid leukemia. *Cancer*. 120, 2130–2141 (2014).

259. I. Radke, M. Götte, C. Kersting, B. Mattsson, L. Kiesel, P. Wülfing, Expression and prognostic impact of the protein tyrosine phosphatases PRL-1, PRL-2, and PRL-3 in breast cancer. *Brit J Cancer*. 95, 347–354 (2006).

260. M. Wei, K. V. Korotkov, J. S. Blackburn, Targeting phosphatases of regenerating liver (PRLs) in cancer. *Pharmacol Therapeut*. 190, 128–138 (2018).

261. A. M. Campbell, Z.-Y. Zhang, Phosphatase of regenerating liver: a novel target for cancer therapy. *Expert Opin Ther Tar.* 18, 555–569 (2014).

262. B. J. Stephens, H. Han, V. Gokhale, D. D. V. Hoff, PRL phosphatases as potential molecular targets in cancer. *Mol Cancer Ther.* 4, 1653–1661 (2005).

263. M. A. Hjort, P. Abdollahi, E. N. Vandsemb, M. H. Fenstad, B. Lund, T. S. Slørdahl, M. Børset, T. B. Rø, Phosphatase of regenerating liver-3 is expressed in acute lymphoblastic leukemia and mediates leukemic cell adhesion, migration and drug resistance. *Oncotarget*. 9, 3549–3561 (2017).

264. K. E. McQueeney, J. M. Salamoun, J. C. Burnett, N. Barabutis, P. Pekic, S. L. Lewandowski, D. C. Llaneza, R. Cornelison, Y. Bai, Z.-Y. Zhang, J. D. Catravas, C. N.

Landen, P. Wipf, J. S. Lazo, E. R. Sharlow, Targeting ovarian cancer and endothelium with an allosteric PTP4A3 phosphatase inhibitor. *Oncotarget*. 9, 8223–8240 (2017).

265. S. Hardy, E. Kostantin, T. Hatzihristidis, Y. Zolotarov, N. Uetani, M. L. Tremblay, Physiological and oncogenic roles of the PRL phosphatases. *Febs J.* 285, 3886–3908 (2018).

266. L. Duciel, L. C. M. Gomez, M. Kondratova, I. Kuperstein, S. Saule, The Phosphatase PRL-3 Is Involved in Key Steps of Cancer Metastasis. *J Mol Biol*. 431, 3056–3067 (2019).

267. K. Guo, J. Li, J. P. Tang, V. Koh, B. Q. Gan, Q. Zeng, Catalytic domain of PRL-3 plays an essential role in tumor metastasis: Formation of PRL-3 tumors inside the blood vessels. *Cancer Biol Ther.* 3, 945–951 (2004).

268. S. Hardy, N. N. Wong, W. J. Muller, M. Park, M. L. Tremblay, Overexpression of the Protein Tyrosine Phosphatase PRL-2 Correlates with Breast Tumor Formation and Progression. *Cancer Res.* 70, 8959–8967 (2010).

269. X. Wu, H. Zeng, X. Zhang, Y. Zhao, H. Sh, X. Ge, M. Zhang, X. Gao, Q. Xu, Phosphatase of Regenerating Liver-3 Promotes Motility and Metastasis of Mouse Melanoma Cells. *Am J Pathol.* 164, 2039–2054 (2010).

270. H. Achiwa, J. S. Lazo, PRL-1 Tyrosine Phosphatase Regulates c-Src Levels, Adherence, and Invasion in Human Lung Cancer Cells. *Cancer Res.* 67, 643–650 (2007).

271. H. Kato, S. Semba, U. A. Miskad, Y. Seo, M. Kasuga, H. Yokozaki, High Expression of PRL-3 Promotes Cancer Cell Motility and Liver Metastasis in Human Colorectal Cancer A Predictive Molecular Marker of Metachronous Liver and Lung Metastases. *Clin Cancer Res.* 10, 7318–7328 (2004).

272. Z. Li, W. Zhan, Z. Wang, B. Zhu, Y. He, J. Peng, S. Cai, J. Ma, Inhibition of PRL-3 gene expression in gastric cancer cell line SGC7901 via microRNA suppressed reduces peritoneal metastasis. *Biochem Bioph Res Co.* 348, 229–237 (2006).

273. F. Polato, A. Codegoni, R. Fruscio, P. Perego, C. Mangioni, S. Saha, A. Bardelli, M. Broggini, PRL-3 Phosphatase Is Implicated in Ovarian Cancer Growth. *Clin Cancer Res.* 11, 6835–6839 (2005).

274. F. Qian, Y.-P. Li, X. Sheng, Z.-C. Zhang, R. Song, W. Dong, S.-X. Cao, Z.-C. Hua, Q. Xu, PRL-3 siRNA Inhibits the Metastasis of B16-BL6 Mouse Melanoma Cells In Vitro and In Vivo. *Mol Med.* 13, 151–159 (2007).

275. V. McParland, G. Varsano, X. Li, J. Thornton, J. Baby, A. Aravind, C. Meyer, K. Pavic, P. Rios, M. Köhn, The Metastasis-Promoting Phosphatase PRL-3 Shows Activity toward Phosphoinositides. *Biochemistry-us.* 50, 7579–7590 (2011).

276. M. Wei, M. G. Haney, D. R. Rivas, J. S. Blackburn, Protein tyrosine phosphatase 4A3 (PTP4A3/PRL-3) drives migration and progression of T-cell acute lymphoblastic leukemia in vitro and in vivo. *Oncogenesis*. 9, 6 (2020).

277. G. Kozlov, J. Cheng, E. Ziomek, D. Banville, K. Gehring, I. Ekiel, Structural Insights into Molecular Function of the Metastasis-associated Phosphatase PRL-3*. *J Biol Chem.* 279, 11882–11889 (2004).

278. K.-A. Kim, J.-S. Song, J. Jee, M. R. Sheen, C. Lee, T. G. Lee, S. Ro, J. M. Cho, W. Lee, T. Yamazaki, Y. H. Jeon, C. Cheong, Structure of human PRL-3, the phosphatase associated with cancer metastasis. *Febs Lett.* 565, 181–187 (2004).

279. K.-W. Jeong, D.-I. Kang, E. Lee, A. Shin, B. Jin, Y.-G. Park, C.-K. Lee, E.-H. Kim, Y. H. Jeon, E. E. Kim, Y. Kim, Structure and Backbone Dynamics of Vanadate-Bound PRL-3: Comparison of 15N Nuclear Magnetic Resonance Relaxation Profiles of Free and Vanadate-Bound PRL-3. *Biochemistry-us.* 53, 4814–4825 (2014).

280. P. Rios, X. Li, M. Köhn, Molecular mechanisms of the PRL phosphatases: Molecular mechanisms of the PRL phosphatases. *Febs J.* 280, 505–524 (2012).

281. J.-P. Sun, W.-Q. Wang, H. Yang, S. Liu, F. Liang, A. A. Fedorov, S. C. Almo, Z.-Y. Zhang, Structure and Biochemical Properties of PRL-1, a Phosphatase Implicated in Cell Growth, Differentiation, and Tumor Invasion †, ‡. *Biochemistry-us*. 44, 12009–12021 (2005).

282. Q. Zeng, X. Si, H. Horstmann, Y. Xu, W. Hong, C. J. Pallen, Prenylation-dependent Association of Protein-tyrosine Phosphatases PRL-1, -2, and -3 with the Plasma Membrane and the Early Endosome*. *J Biol Chem.* 275, 21444–21452 (2000).

283. J. N. Andersen, O. H. Mortensen, G. H. Peters, P. G. Drake, L. F. Iversen, O. H. Olsen, P. G. Jansen, H. S. Andersen, N. K. Tonks, N. P. H. Møller, Structural and Evolutionary Relationships among Protein Tyrosine Phosphatase Domains. *Mol Cell Biol.* 21, 7117–7136 (2001).

284. S. Daouti, W. Li, H. Qian, K.-S. Huang, J. Holmgren, W. Levin, L. Reik, D. L. McGady, P. Gillespie, A. Perrotta, H. Bian, J. F. Reidhaar-Olson, S. A. Bliss, A. R. Olivier, J. A. Sergi, D. Fry, W. Danho, S. Ritland, N. Fotouhi, D. Heimbrook, H. Niu, A Selective Phosphatase of Regenerating Liver Phosphatase Inhibitor Suppresses Tumor Cell Anchorage-Independent Growth by a Novel Mechanism Involving p130Cas Cleavage. *Cancer Res.* 68, 1162–1169 (2008).

285. J. M. Salamoun, K. E. McQueeney, K. Patil, S. J. Geib, E. R. Sharlow, J. S. Lazo, P. Wipf, Photooxygenation of an amino-thienopyridone yields a more potent PTP4A3 inhibitor. *Org Biomol Chem.* 14, 6398–6402 (2016).

286. G. Min, S.-K. Lee, H.-N. Kim, Y.-M. Han, R.-H. Lee, D. G. Jeong, D. C. Han, B.-M. Kwon, Rhodanine-based PRL-3 inhibitors blocked the migration and invasion of metastatic cancer cells. *Bioorg Med Chem Lett.* 23, 3769–3774 (2013).

287. B. Hoeger, M. Diether, P. J. Ballester, M. Köhn, Biochemical evaluation of virtual screening methods reveals a cell-active inhibitor of the cancer-promoting phosphatases of regenerating liver. *Eur J Med Chem.* 88, 89–100 (2014).

288. M. K. Pathak, D. Dhawan, D. J. Lindner, E. C. Borden, C. Farver, T. Yi, Pentamidine is an inhibitor of PRL phosphatases with anticancer activity. *Mol Cancer Ther.* 1, 1255–64 (2002).

289. S.-K. Choi, H.-M. Oh, S.-K. Lee, D. G. Jeong, S. E. Ryu, K.-H. Son, D. C. Han, N.-D. Sung, N.-I. Baek, B.-M. Kwon, Biflavonoids inhibited phosphatase of regenerating liver-3 (PRL-3). *Nat Prod Res.* 20, 341–346 (2006).

290. L. Wang, Y. Shen, R. Song, Y. Sun, J. Xu, Q. Xu, An Anticancer Effect of Curcumin Mediated by Down-Regulating Phosphatase of Regenerating Liver-3 Expression on Highly Metastatic Melanoma Cells. *Mol Pharmacol.* 76, 1238–1245 (2009).

291. Y.-M. Han, S.-K. Lee, D. G. Jeong, S. E. Ryu, D. C. Han, D. K. Kim, B.-M. Kwon, Emodin inhibits migration and invasion of DLD-1 (PRL-3) cells via inhibition of PRL-3 phosphatase activity. *Bioorg Med Chem Lett.* 22, 323–326 (2012).

292. Y. Bai, Z.-H. Yu, S. Liu, L. Zhang, R.-Y. Zhang, L.-F. Zeng, S. Zhang, Z.-Y. Zhang, Novel Anticancer Agents Based on Targeting the Trimer Interface of the PRL Phosphatase. *Cancer Res.* 76, 4805–4815 (2016).

293. Y. Bai, Z.-H. Yu, Z.-Y. Zhang, Discovery and Evaluation of PRL Trimer Disruptors for Novel Anticancer Agents. *Methods Mol Biology Clifton N J.* 1447, 121–38 (2016).

294. D. G. Jeong, S. J. Kim, J. H. Kim, J. H. Son, M. R. Park, S. M. Lim, T.-S. Yoon, S. E. Ryu, Trimeric Structure of PRL-1 Phosphatase Reveals an Active Enzyme Conformation and Regulation Mechanisms. *J Mol Biol.* 345, 401–413 (2005).

295. D. Lee, C. Hilty, G. Wider, K. Wüthrich, Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson*. 178, 72–76 (2006).

296. M. K. Pathak, D. Dhawan, D. J. Lindner, E. C. Borden, C. Farver, T. Yi, Pentamidine Is an Inhibitor of PRL Phosphatases with Anticancer Activity. *Mol Cancer Ther.* 1, 1255–1264 (2002).

297. Z. Zhang, G. Kozlov, Y. S. Chen, K. Gehring, Mechanism of thienopyridone and iminothienopyridinedione inhibition of protein phosphatases. *Medchemcomm*. 10, 791–799 (2019).

298. D. Parisi, M. F. Adasme, A. Sveshnikova, S. N. Bolz, Y. Moreau, M. Schroeder, Drug repositioning or target repositioning: A structural perspective of drug-targetindication relationship for available repurposed drugs. *Comput Struct Biotechnology J.* 18, 1043–1055 (2020).

299. S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, M. Pirmohamed, Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov.* 18, 41–58 (2019).

300. R. Kumar, S. Harilal, S. V. Gupta, J. Jose, D. G. T. Parambi, Md. S. Uddin, M. A. Shah, B. Mathew, Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work. *Eur J Med Chem.* 182, 111602 (2019).

301. T. I. Oprea, J. Mestres, Drug Repurposing: Far Beyond New Targets for Old Drugs. *Aaps J.* 14, 759–763 (2012).

302. E. Gregori-Puigjané, V. Setola, J. Hert, B. Crews, J. Irwin, E. Lounkine, L. Marnett, B. Roth, B. Shoichet, Identifying mechanism-of-action targets for drugs and probes. *PNAS*. 109, 11178–11183 (2012).

303. S. M. Corsello, J. A. Bittker, Z. Liu, J. Gould, P. McCarren, J. E. Hirschman, S. E. Johnston, A. Vrcic, B. Wong, M. Khan, J. Asiedu, R. Narayan, C. C. Mader, A. Subramanian, T. R. Golub, The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med.* 23, 405–408 (2017).

304. D. R. Rivas, M. V. C. D. Cerna, C. N. Smith, S. Sampathi, B. G. Patty, D. Lee, J. S. Blackburn, A screen of FDA-approved drugs identifies inhibitors of protein tyrosine phosphatase 4A3 (PTP4A3 or PRL-3). *Sci Rep-uk*. 11, 10302 (2021).

305. D. Sladowski, S. J. Steer, R. H. Clothier, M. Balls, An improved MIT assay. J Immunol Methods. 157, 203–207 (1993).

306. C.-C. Liang, A. Y. Park, J.-L. Guan, In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat Protoc.* 2, 329–333 (2007).

307. M. Gana-Weisz, R. Haklai, D. Marciano, Y. Egozi, G. Ben-Baruch, Y. Kloog, The Ras AntagonistS-Farnesylthiosalicylic Acid Induces Inhibition of MAPK Activation. *Biochem Bioph Res Co.* 239, 900–904 (1997).

308. G. Elad, A. Paz, R. Haklai, D. Marciano, A. Cox, Y. Kloog, Targeting of K-Ras 4B by S-trans, trans-farnesyl thiosalicylic acid. *Biochimica Et Biophysica Acta Bba - Mol Cell Res.* 1452, 228–242 (1999).

309. C. A. Rowell, J. J. Kowalczyk, M. D. Lewis, A. M. Garcia, Direct Demonstration of Geranylgeranylation and Farnesylation of Ki-Ras in Vivo *. *J Biol Chem*. 272, 14093–14097 (1997).

310. T. Rubio, M. Köhn, Regulatory mechanisms of phosphatase of regenerating liver (PRL)-3. *Biochem Soc T*. 44, 1305–1312 (2016).

311. M. Pascaru, C. Tanase, A. M. Vacaru, P. Boeti, E. Neagu, I. Popescu, S. E. Szedlacsek, Analysis of molecular determinants of PRL-3. *J Cell Mol Med.* 13, 3141–3150 (2009).

312. T. UNGER, Differences Among Angiotensin II Type 1 Receptor Blockers: Characteristics of Candesartan Cilexetil. *Blood Pressure*. 9, 14–18 (2000).

313. Y. Inada, M. Ojima, R. Kanagawa, Y. Misumi, K. Nishikawa, T. Naka, Pharmacologic properties of candesartan cilexetil—possible mechanisms of long-acting antihypertensive action. *J Hum Hypertens*. 13, S75–S80 (1999).

314. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 30, 2785–2791 (2009).

315. Q. Li, T. Cheng, Y. Wang, S. H. Bryant, PubChem as a public resource for drug discovery. *Drug Discov Today*. 15, 1052–1057 (2010).

316. Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, S. H. Bryant, PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37, W623–W633 (2009).

317. M. M. Jaghoori, B. Bleijlevens, S. D. Olabarriaga, 1001 Ways to run AutoDock Vina for virtual screening. *J Comput Aid Mol Des.* 30, 237–249 (2016).

318. A. Luttens, H. Gullberg, E. Abdurakhmanov, D. D. Vo, D. Akaberi, V. O. Talibov, N. Nekhotiaeva, L. Vangeel, S. D. Jonghe, D. Jochmans, J. Krambrich, A. Tas, B. Lundgren, Y. Gravenfors, A. J. Craig, Y. Atilaw, A. Sandström, L. W. K. Moodie, Å. Lundkvist, M. J. van Hemert, J. Neyts, J. Lennerstrand, J. Kihlberg, K. Sandberg, U. H. Danielson, J. Carlsson, Ultralarge Virtual Screening Identifies SARS-CoV-2 Main Protease Inhibitors with Broad-Spectrum Activity against Coronaviruses. *J Am Chem Soc.* 144, 2905–2920 (2022).

319. C. Gorgulla, A. Boeszoermenyi, Z.-F. Wang, P. D. Fischer, P. W. Coote, K. M. P. Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner, H. Arthanari, An open-source drug discovery platform enables ultra-large virtual screens. *Nature*. 580, 663–668 (2020).

320. T. M. Bobrowski, D. R. Korn, E. N. Muratov, A. Tropsha, ZINC Express: A Virtual Assistant for Purchasing Compounds Annotated in the ZINC Database. *J Chem Inf Model*. 61, 1033–1036 (2021).

321. J. Irwin, B. Shiochet, ZINC - a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*. 45, 177–182 (2005).

322. M. Congreve, C. W. Murray, T. L. Blundell, Keynote review: Structural biology and drug discovery. *Drug Discov Today*. 10, 895–907 (2005).

323. W. S. Selby, G. D. Barr, A. Ireland, C. H. Mason, D. P. Jewell, Olsalazine in active ulcerative colitis. *Br Medical J Clin Res Ed*. 291, 1373 (1985).

324. M. G. Courtney, D. P. Nunes, C. F. Bergin, M. O'Driscoll, V. Trimble, P. W. N. Keeling, D. G. Weir, Randomised comparison of olsalazine and mesalazine in prevention of relapses in ulcerative colitis. *Lancet*. 339, 1279–1281 (1992).

325. T. Sousa, Y. Yadav, V. Zann, A. Borde, B. Abrahamsson, A. Basit, On the colonic bacterial metabolism of azo-bonded prodrugs of 5-aminosalicylic acid. *Journal of Pharmaceutical Sciences*. 103, 3171–3175 (2014).

326. M. N. Schulz, J. Landström, K. Bright, R. E. Hubbard, Design of a Fragment Library that maximally represents available chemical space. *J Comput Aid Mol Des.* 25, 611 (2011).

327. H. Zhang, G. Kozlov, X. Li, H. Wu, I. Gulerez, K. Gehring, PRL3 phosphatase active site is required for binding the putative magnesium transporter CNNM3. *Sci Rep-uk.* 7, 48 (2017).

328. S. K. Whittier, A. C. Hengge, J. P. Loria, Conformational Motions Regulate Phosphoryl Transfer in Related Protein Tyrosine Phosphatases. *Science*. 341, 899–903 (2013).

329. S. C. L. Kamerlin, R. Rucker, S. Boresch, A molecular dynamics study of WPD-loop flexibility in PTP1B. *Biochem Bioph Res Co.* 356, 1011–1016 (2007).

330. T. A. S. Brandão, S. J. Johnson, A. C. Hengge, The molecular details of WPD-loop movement differ in the protein-tyrosine phosphatases YopH and PTP1B. *Arch Biochem Biophys.* 525, 53–59 (2012).

331. J. Huang, A. D. MacKerell, CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem.* 34, 2135–2145 (2013).

332. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD. *J Comput Chem.* 26, 1781–1802 (2005).

333. H. J. C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A messagepassing parallel molecular dynamics implementation. *Comput Phys Commun.* 91, 43–56 (1995).

334. J. D. Durrant, J. A. McCammon, Molecular dynamics simulations and drug discovery. *Bmc Biol.* 9, 71–71 (2011).

335. G. Kar, O. Keskin, A. Gursoy, R. Nussinov, Allostery and population shift in drug discovery. *Curr Opin Pharmacol*. 10, 715–722 (2010).

336. D. D. Boehr, J. R. Schnell, D. McElheny, S.-H. Bae, B. M. Duggan, S. J. Benkovic, H. J. Dyson, P. E. Wright, A Distal Mutation Perturbs Dynamic Amino Acid Networks in Dihydrofolate Reductase. *Biochemistry-us.* 52, 4605–4619 (2013).

337. X. Peng, C. Lu, J. Pang, Z. Liu, D. Lu, A distal regulatory strategy of enzymes: from local to global conformational dynamics. *Phys Chem Chem Phys.* 23, 22451–22465 (2021).

338. P. Campitelli, T. Modi, S. Kumar, S. B. Ozkan, The role of conformational dynamics and allostery in modulating protein evolution. *Annual Review of Biophysics*. 49, 267–288 (2020).

339. H. M. Dingerdissen, J. Torcivia-Rodriguez, Y. Hu, T.-C. Chang, R. Mazumder, R. Kahsay, BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. *Nucleic Acids Res.* 46, D1128–D1136 (2018).

340. N. Lyle, R. K. Das, R. V. Pappu, A quantitative measure for protein conformational heterogeneity. *J Chem Phys.* 139, 121907 (2013).

341. D. P. LANE, L. V. CRAWFORD, T antigen is bound to a host protein in SY40-transformed cells. *Nature*. 278, 261–263 (1979).

342. C. A. Finlay, P. W. Hinds, A. J. Levine, The p53 proto-oncogene can act as a suppressor of transformation. *Cell*. 57, 1083–1093 (1989).

343. D. P. Lane, p53, guardian of the genome. Nature. 358, 15-16 (1992).

344. U. K. Mukhopadhyay, A. S. Mak, p53: Guardian of the genome is also a suppressor of cell invasion? *Cell Cycle*. 8, 2481–2481 (2009).

345. M. Gomez-Lazaro, F. J. Fernandez-Gomez, J. Jordán, p53: Twenty five years understanding the mechanism of genome protection. *J Physiol Biochem*. 60, 287–307 (2004).

346. E. Toufektchan, F. Toledo, The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. *Cancers*. 10, 135 (2018).

347. S. Bursac, M. C. Brdovcak, G. Donati, S. Volarevic, Activation of the tumor suppressor p53 upon impairment of ribosome biogenesis. *Biochimica Et Biophysica Acta Bba - Mol Basis Dis.* 1842, 817–830 (2014).

348. L. Golomb, S. Volarevic, M. Oren, p53 and ribosome biogenesis stress: The essentials. *Febs Lett.* 588, 2571–2579 (2014).

349. S. W. Lowe, Activation of p53 by oncogenes. *Endocr-relat Cancer*. 6, 45–48 (1999).

350. S. Y. Fuchs, V. Adler, T. Buschmann, Z. Yin, X. Wu, S. N. Jones, Z. Ronai, JNK targets p53 ubiquitination and degradation in nonstressed cells. *Gene Dev.* 12, 2658–2663 (1998).

351. S. Y. Fuchs, V. Adler, M. R. Pincus, Z. Ronai, MEKK1/JNK signaling stabilizes and activates p53. *Proc National Acad Sci*. 95, 10541–10546 (1998).

352. D. V. Bulavin, S. Saito, M. C. Hollander, K. Sakaguchi, C. W. Anderson, E. Appella, A. J. Fornace, Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *Embo J.* 18, 6845–6854 (1999).

353. S. L. Harris, A. J. Levine, The p53 pathway: positive and negative feedback loops. *Oncogene*. 24, 2899–2908 (2005).

354. S. Ingvarsson, B. I. Sigbjornsdottir, C. Huiping, S. H. Hafsteinsdottir, G. Ragnarsson, R. B. Barkardottir, A. Arason, V. Egilsson, J. T. Bergthorsson, Mutation analysis of the CHK2 gene in breast carcinoma and other cancers. *Breast Cancer Res.* 4, R4–R4 (2002).

355. J. Bartek, J. Lukas, Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell.* 3, 421–429 (2003).

356. T. Buschmann, O. Potapova, A. Bar-Shira, V. Ivanov, S. Fuchs, S. Henderson, V. Fried, T. Minamoto, D. Alarcon-Vargas, M. Pincus, W. Gaarde, N. Holbrook, Y. Shiloh, Z. Ronai, Jun NH2-Terminal Kinase Phosphorylation of p53 on Thr-81 Is Important for p53 Stabilization and Transcriptional Activities in Response to Stress. *Molecular and Cellular Biology*. 21, 2743–2754 (2001).

357. M. Hollstein, D. Sidransky, B. Vogelstein, C. Harris, p53 mutations in human cancers. *Science*. 253, 49–53 (1991).

358. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types. *Nature*. 502, 333–339 (2013).

359. R. Brosh, V. Rotter, When mutants gain new powers: news from the mutant p53 field. *Nat Rev Cancer*. 9, 701–713 (2009).

360. D. Dittmer, S. Pati, G. Zambetti, S. Chu, A. K. Teresky, M. Moore, C. Finlay, A. J. Levine, Gain of function mutations in p53. *Nat Genet.* 4, 42–46 (1993).

361. G. Shaulsky, N. Goldfinger, V. Rotter, Alterations in tumor development in vivo mediated by expression of wild type or mutant p53 proteins. *Cancer Res.* 51, 5232–7 (1991).

362. D. Wolf, N. Harris, V. Rotter, Reconstitution of p53 expression in a nonproducer Ab-MuLV-transformed cell line by transfection of a functional p53 gene. *Cell*. 38, 119–126 (1984).

363. V. J. N. Bykov, S. E. Eriksson, J. Bianchi, K. G. Wiman, Targeting mutant p53 for efficient cancer therapy. *Nat Rev Cancer*. 18, 89–102 (2018).

364. G. Zhu, C. Pan, J.-X. Bei, B. Li, C. Liang, Y. Xu, X. Fu, Mutant p53 in Cancer Progression and Targeted Therapies. *Frontiers Oncol.* 10, 595187 (2020).

365. D. A. Carson, A. Lois, Cancer progression and p53. Lancet. 346, 1009–1011 (1995).

366. J. Momand, G. P. Zambetti, D. C. Olson, D. George, A. J. Levine, The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell*. 69, 1237–1245 (1992).

367. Y. Zhao, H. Yu, W. Hu, The regulation of MDM2 oncogene and its impact on human cancers. *Acta Bioch Bioph Sin.* 46, 180–189 (2014).

368. X. Wu, J. H. Bayle, D. Olson, A. J. Levine, The p53-mdm-2 autoregulatory feedback loop. *Gene Dev.* 7, 1126–1132 (1993).

369. A. Shvarts, W. T. Steegenga, N. Riteco, T. Laar, P. Dekker, M. Bazuine, R. C. Ham, W. H. van Oordt, G. Hateboer, A. J. Eb, A. G. Jochemsen, MDMX: a novel p53-binding protein with some functional properties of MDM2. *Embo J.* 15, 5349–5357 (1996).

370. A. Shvarts, M. Bazuine, P. Dekker, Y. Ramos, W. Steegenga, G. Merckx, R. van Ham, W. van der H. van Oordt, A. van der Eb, A. Jochemsen, Isolation and identification

of the human homolog of a new p53-binding protein, mdmx. *Genomics*. 43, 34–42 (1997).

371. S. Ménendez, A. M. Goh, S. Camus, K. W. Ng, N. Kua, V. Badal, D. P. Lane, MDM4 downregulates p53 transcriptional activity and response to stress during differentiation. *Cell Cycle*. 10, 1100–1108 (2011).

372. T. Iwakuma, G. Lozano, MDM2: an introduction. *Molecular Cancer Research*. 1, 993–1000 (2003).

373. M. Shadfan, V. Lopez-Pajares, Z.-M. Yuan, MDM2 and MDMX: alone and together in regulation of p53. *Translational Cancer Research*. 1, 88–99 (2012).

374. D. Pei, Y. Zhang, J. Zheng, Regulation of p53: a collaboration between Mdm2 and MdmX. *Oncotarget.* 3, 228–235 (2012).

375. J.-C. Marine, A. G. Jochemsen, Mdmx and Mdm2: Brothers in Arms? *Cell Cycle*. 3, 898–902 (2004).

376. C. L. Brooks, W. Gu, p53 Ubiquitination: Mdm2 and Beyond. *Mol Cell*. 21, 307–315 (2006).

377. G. M. Popowicz, A. Czarna, U. Rothweiler, A. Szwagierczak, M. Krajewski, L. Weber, T. A. Holak, Molecular Basis for the Inhibition of p53 by Mdmx. *Cell Cycle*. 6, 2386–2392 (2007).

378. L. K. Linares, A. Hengstermann, A. Ciechanover, S. Müller, M. Scheffner, HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53. *Proc National Acad Sci.* 100, 12009–12014 (2003).

379. J. Gu, H. Kawai, L. Nie, H. Kitao, D. Wiederschain, A. G. Jochemsen, J. Parant, G. Lozano, Z.-M. Yuan, Mutual Dependence of MDM2 and MDMX in Their Functional Inactivation of p53*. *J Biol Chem.* 277, 19251–19254 (2002).

380. V. Böttger, A. Böttger, C. Garcia-Echeverria, Y. F. Ramos, A. J. van der Eb, A. G. Jochemsen, D. P. Lane, Comparative study of the p53-mdm2 and p53-MDMX interfaces. *Oncogene*. 18, 189–199 (1999).

381. M. Wade, G. M. Wahl, Targeting Mdm2 and Mdmx in Cancer Therapy: Better Living through Medicinal Chemistry? *Mol Cancer Res.* 7, 1–11 (2009).

382. Y. Zhao, D. Bernard, S. Wang, Small Molecule Inhibitors of MDM2-p53 and MDMX-p53 Interactions as New Cancer Therapeutics. *Biodiscovery*. 8, e8950 (2013).

383. C. R. Grace, D. Ban, J. Min, A. Mayasundari, L. Min, K. E. Finch, L. Griffiths, N. Bharatham, D. Bashford, R. K. Guy, M. A. Dyer, R. W. Kriwacki, Monitoring Ligand-Induced Protein Ordering in Drug Discovery. *J Mol Biol.* 428, 1290–1303 (2016).

384. W. Borcherds, F.-X. Theillet, A. Katzer, A. Finzel, K. M. Mishall, A. T. Powell, H. Wu, W. Manieri, C. Dieterich, P. Selenko, A. Loewer, G. W. Daughdrill, Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat Chem Biol.* 10, 1000–1002 (2014).

385. G. M. Popowicz, A. Dömling, T. A. Holak, The Structure-Based Design of Mdm2/Mdmx-p53 Inhibitors Gets Serious. *Angewandte Chemie Int Ed.* 50, 2680–2688 (2011).

386. B. Anil, C. Riedinger, J. A. Endicott, M. E. M. Noble, The structure of an MDM2– Nutlin-3a complex solved by the use of a validated MDM2 surface-entropy reduction mutant. *Acta Crystallogr Sect D Biological Crystallogr*. 69, 1358–1366 (2013).

387. G. Popowicz, A. Czarna, T. Holak, Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. *Cell Cycle*. 7, 2441–2443 (2008).

388. A. Ramanathan, A. Savol, V. Burger, C. S. Chennubhotla, P. K. Agarwal, Protein Conformational Populations and Functionally Relevant Substates. *Accounts Chem Res.* 47, 149–156 (2014).

389. S. Grutsch, S. Brüschweiler, M. Tollinger, NMR Methods to Study Dynamic Allostery. *Plos Comput Biol.* 12, e1004620 (2016).

390. A. G. Palmer, F. Massi, Characterization of the Dynamics of Biomacromolecules Using Rotating-Frame Spin Relaxation NMR Spectroscopy. *Chem Rev.* 106, 1700–1719 (2006).

391. L. E. Kay, New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J Mol Biol.* 428, 323–331 (2016).

392. G. Popowicz, A. Czarna, T. Holak, Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. *Cell Cycle*. 7, 2441–2443 (2008).

393. B. Graves, T. Thompson, M. Xia, C. Janson, C. Lukacs, D. Deo, P. D. Lello, D. Fry, C. Garvie, K.-S. Huang, L. Gao, C. Tovar, A. Lovey, J. Wanner, L. T. Vassilev, Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization. *Proc National Acad Sci.* 109, 11788–11793 (2012).

394. M. Zweckstetter, A. Bax, Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J Biomol Nmr*. 20, 365–377 (2001).

395. M. R. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol.* 5, 1065–1074 (1998).

396. J. G. Reddy, S. Pratihar, D. Ban, S. Frischkorn, S. Becker, C. Griesinger, D. Lee, Simultaneous determination of fast and slow dynamics in molecules using extreme CPMG relaxation dispersion experiments. *J Biomol Nmr*. 70, 1–9 (2018).

397. C. A. Smith, D. Ban, S. Pratihar, K. Giller, M. Paulat, S. Becker, C. Griesinger, D. Lee, B. L. de Groot, Allosteric switch regulates protein–protein binding through collective motion. *Proc National Acad Sci.* 113, 3269–3274 (2016).

398. S. Eriksson, B. Munch-Petersen, K. Johansson, H. Ecklund, Structure and function of cellular deoxyribonucleoside kinases. *Cell Mol Life Sci Cmls*. 59, 1327–1346 (2002).

399. A. R. V. Rompay, M. Johansson, A. Karlsson, Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases. *Pharmacol Therapeut*. 100, 119–139 (2003).

400. P. Reichard, Interactions between deoxyribonucleotide and DNA synthesis. *Annual Review of Biochemistry*. 57, 349–374 (1988).

401. A. R. V. Rompay, M. Johansson, A. Karlsson, Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases. *Pharmacol Therapeut*. 87, 189–198 (2000).

402. D. Gentry, C. Bengra, K. Ikehara, M. Cashel, Guanylate kinase of Escherichia coli K-12. *J Biol Chem.* 268, 14316–14321 (1993).

403. M. Konrad, Cloning and expression of the essential gene for guanylate kinase from yeast. *Journal of Biological Chemistry*. 267, 25652–25655 (1992).

404. W. A. Brady, M. S. Kokoris, M. Fitzgibbon, M. E. Black, Cloning, Characterization, and Modeling of Mouse and Human Guanylate Kinases*. *J Biol Chem.* 271, 16734–16740 (1996).

405. T. Stehle, G. E. Schulz, Refined structure of the complex between guanylate kinase and its substrate GMP at 2.0 A resolution. *J Mol Biol*. 224, 1127–41 (1992).

406. N. G. Johansson, S. Eriksson, Structure-activity relationships for phosphorylation of nucleoside analogs to monophosphates by nucleoside kinases. *Acta Biochim Pol.* 43, 143–60 (1996).

407. B. G. Gentry, S. N. Gentry, T. L. Jackson, J. Zemlicka, J. C. Drach, Phosphorylation of antiviral and endogenous nucleotides to di- and triphosphates by guanosine monophosphate kinase. *Biochem Pharmacol.* 81, 43–49 (2011).

408. W. T. Ashton, J. D. Karkas, A. K. Field, R. L. Tolman, Activation by thymidine kinase and potent antiherpetic activity of 2'-nor-2'-deoxyguanosine (2'NDG). *Biochem Bioph Res Co.* 108, 1716–1721 (1982).

409. D. Deville-Bonne, C. E. Amri, P. Meyer, Y. Chen, L. A. Agrofoglio, J. Janin, Human and viral nucleoside/nucleotide kinases involved in antiviral drug activation: Structural and catalytic properties. *Antivir Res.* 86, 101–120 (2010).

410. R. L. Miller, D. L. Adamczyk, T. Spector, K. C. Agarwal, R. P. Miech, R. E. Parks, Reassessment of the interactions of guanylate kinase and 6-thioguanosine 5'-phosphate. *Biochem Pharmacol.* 26, 1573–1576 (1977).

411. T. W. Traut, Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem.* 140, 1–22 (1994).

412. O. Shuvalov, A. Petukhov, A. Daks, O. Fedorova, E. Vasileva, N. A. Barlev, Onecarbon metabolism and nucleotide biosynthesis as attractive targets for anticancer therapy. *Oncotarget*. 8, 23955–23977 (2017).

413. N. Majd, K. Sumita, H. Yoshino, D. Chen, J. Terakawa, T. Daikoku, S. Kofuji, R. Curry, T. M. Wise-Draper, R. E. Warnick, J. Guarnaschelli, A. T. Sasaki, A Review of the Potential Utility of Mycophenolate Mofetil as a Cancer Therapeutic. *J Cancer Res.* 2014, 1–12 (2014).

414. E. Oláh, S. Kökény, J. Papp, A. Bozsik, M. Keszei, Modulation of cancer pathways by inhibitors of guanylate metabolism. *Adv Enzyme Regul.* 46, 176–190 (2006).

415. S. Kofuji, A. Sasaki, GTP metabolic reprograming by IMPDH2: unlocking cancer cells' fuelling mechanism. *J Biochem*. 168, 319–328 (2020).

416. Q. Shu, V. Nair, Inosine monophosphate dehydrogenase (IMPDH) as a target in drug discovery. *Med Res Rev.* 28, 219–232 (2008).

417. Q. Wang, Y. F. Guan, S. E. Hancock, K. Wahi, M. van Geldermalsen, B. K. Zhang, A. Pang, R. Nagarajah, B. Mak, N. Freidman, L. G. Horvath, N. Turner, J. Holst, Inhibition of guanosine monophosphate synthetase (GMPS) blocks glutamine metabolism and prostate cancer growth. *J Pathology*. 254, 135–146 (2021).

418. R. Naffouje, P. Grover, H. Yu, A. Sendilnathan, K. Wolfe, N. Majd, E. P. Smith, K. Takeuchi, T. Senda, S. Kofuji, A. T. Sasaki, Anti-Tumor Potential of IMP Dehydrogenase Inhibitors: A Century-Long Story. *Cancers*. 11, 1346 (2019).

419. L. Hedstrom, IMP Dehydrogenase: Structure, Mechanism, and Inhibition. *Chem Rev.* 109, 2903–2928 (2009).

420. A. Bianchi-Smiraglia, J. A. Wawrzyniak, A. Bagati, E. K. Marvin, J. Ackroyd, S. Moparthy, W. Bshara, E. E. Fink, C. E. Foley, G. E. Morozevich, A. E. Berman, D. S. Shewach, M. A. Nikiforov, Pharmacological targeting of guanosine monophosphate synthase suppresses melanoma cell invasion and tumorigenicity. *Cell Death Differ*. 22, 1858–1864 (2015).

421. N. Khan, D. Ban, P. Trigo-Mourino, M. Carneiro, M. Konrad, D. Lee, T. M. Sabo, 1H, 13C, and 15N resonance assignment of human guanylate kinase. *Biomolecular NMR Assignments*. 12, 11–14 (2018).

422. G. Hible, L. Renault, F. Schaeffer, P. Christova, A. Z. Radulescu, C. Evrin, A.-M. Gilles, J. Cherfils, Calorimetric and Crystallographic Analysis of the Oligomeric Structure of Escherichia coli GMP Kinase. *J Mol Biol.* 352, 1044–1059 (2005).

423. Y. Li, Y. Zhang, H. Yan, Kinetic and Thermodynamic Characterizations of Yeast Guanylate Kinase*. *J Biol Chem*. 271, 28038–28044 (1996).

424. Y. Zhang, H. Niu, Y. Li, H. Chu, H. Shen, D. Zhang, G. Li, Mechanistic insight into the functional transition of the enzyme guanylate kinase induced by a single mutation. *Sci Rep-uk.* 5, 8405 (2015).

425. O. Delalande, S. Sacquin-Mora, M. Baaden, Enzyme Closure and Nucleotide Binding Structurally Lock Guanylate Kinase. *Biophys J.* 101, 1440–1449 (2011).

426. N. Sekulic, L. Shuvalova, O. Spangenberg, M. Konrad, A. Lavie, Structural Characterization of the Closed Conformation of Mouse Guanylate Kinase*. *J Biol Chem.* 277, 30236–30243 (2002).

427. R. Jain, N. Khan, A. Menzel, I. Rajkovic, M. Konrad, S. Techert, Insights into open/closed conformations of the catalytically active human guanylate kinase as investigated by small-angle X-ray scattering. *Eur Biophys J.* 45, 81–89 (2016).

428. J. Blaszczyk, Y. Li, H. Yan, X. Ji, Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes11Edited by I. Wilson. *J Mol Biol.* 307, 247–257 (2001).

429. A. G. Palmer, Probing molecular motion by NMR. *Curr Opin Struc Biol*. 7, 732–737 (1997).

430. O. Delalande, S. Sacquin-Mora, M. Baaden, Enzyme Closure and Nucleotide Binding Structurally Lock Guanylate Kinase. *Biophys J.* 101, 1440–1449 (2011).

431. A. G. Palmer, Enzyme Dynamics from NMR Spectroscopy. *Accounts Chem Res.* 48, 457–465 (2015).

432. M. Akke, NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Curr Opin Struc Biol.* 12, 642–647 (2002).

433. D. S. Whitney, B. F. Volkman, K. E. Prehoda, Evolution of a Protein Interaction Domain Family by Tuning Conformational Flexibility. *J Am Chem Soc.* 138, 15150–15156 (2016).

434. K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, D. Kern, Intrinsic motions along an enzymatic reaction trajectory. *Nature*. 450, 838–844 (2007).

435. Z. Luz, S. Meiboom, Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *J Chem Phys.* 39, 366–370 (1963).

436. V. V. Sakhrani, R. K. Ghosh, E. Hilario, K. L. Weiss, L. Coates, L. J. Mueller, Toho-1 β -lactamase: backbone chemical shift assignments and changes in dynamics upon binding with avibactam. *J Biomol Nmr*. 75, 303–318 (2021).

437. W. Elings, A. Gaur, A. Blok, M. Timmer, H. van Ingen, M. Ubink, β -lactamase of Mycobacterium tuberculosis shows dynamics in the active site that increases upon inhibitor binding. *Antimicrobial Agents and Chemotherapy*. 64, 1–13 (2020).

438. C. Müller, G. Schlauderer, J. Reinstein, G. Schulz, Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure*. 4, 147–156 (1996).

439. A. W. Fischer, E. Bordignon, S. Bleicken, A. J. García-Sáez, G. Jeschke, J. Meiler, Pushing the size limit of de novo structure ensemble prediction guided by sparse SDSL-EPR restraints to 200 residues: The monomeric and homodimeric forms of BAX. *J Struct Biol.* 195, 62–71 (2016).

440. P. Cramer, AlphaFold2 and the future of structural biology. *Nat Struct Mol Biol*. 28, 704–705 (2021).

441. R. J. Ellis, Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci.* 26, 597–604 (2001).

442. N. Altincekic, S. M. Korn, N. S. Qureshi, M. Dujardin, M. Ninot-Pedrosa, R. Abele, M. J. A. Saad, C. Alfano, F. C. L. Almeida, I. Alshamleh, G. C. de Amorim, T. K. Anderson, C. D. Anobom, C. Anorma, J. K. Bains, A. Bax, M. Blackledge, J. Blechar, A. Böckmann, L. Brigandat, A. Bula, M. Bütikofer, A. R. Camacho-Zarco, T. Carlomagno, I. P. Caruso, B. Ceylan, A. Chaikuad, F. Chu, L. Cole, M. G. Crosby, V. de Jesus, K.

Dhamotharan, I. C. Felli, J. Ferner, Y. Fleischmann, M.-L. Fogeron, N. K. Fourkiotis, C. Fuks, B. Fürtig, A. Gallo, S. L. Gande, J. A. Gerez, D. Ghosh, F. Gomes-Neto, O. Gorbatyuk, S. Guseva, C. Hacker, S. Häfner, B. Hao, B. Hargittay, K. Henzler-Wildman, J. C. Hoch, K. F. Hohmann, M. T. Hutchison, K. Jaudzems, K. Jović, J. Kaderli, G. Kalniņš, I. Kaņepe, R. N. Kirchdoerfer, J. Kirkpatrick, S. Knapp, R. Krishnathas, F. Kutz, S. zur Lage, R. Lambertz, A. Lang, D. Laurents, L. Lecoq, V. Linhard, F. Löhr, A. Malki, L. M. Bessa, R. W. Martin, T. Matzel, D. Maurin, S. W. McNutt, N. C. Mebus-Antunes, B. H. Meier, N. Meiser, M. Mompeán, E. Monaca, R. Montserret, L. M. Perez, C. Moser, C. Muhle-Goll, T. C. Neves-Martins, X. Ni, B. Norton-Baker, R. Pierattelli, L. Pontoriero, Y. Pustovalova, O. Ohlenschläger, J. Orts, A. T. D. Poian, D. J. Pyper, C. Richter, R. Riek, C. M. Rienstra, A. Robertson, A. S. Pinheiro, R. Sabbatella, N. Salvi, K. Saxena, L. Schulte, M. Schiavina, H. Schwalbe, M. Silber, M. da S. Almeida, M. A. Sprague-Piercy, G. A. Spyroulias, S. Sreeramulu, J.-N. Tants, K. Tārs, F. Torres, S. Töws, M. Á. Treviño, S. Trucks, A. C. Tsika, K. Varga, Y. Wang, M. E. Weber, J. E. Weigand, C. Wiedemann, J. Wirmer-Bartoschek, M. A. W. Martin, J. Zehnder, M. Hengesbach, A. Schlundt, Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. Frontiers Mol Biosci. 8, 653148 (2021).

443. M. A. Danielson, J. J. Falke, Use of 19F NMR to Probe Protein Structure and Conformational Changes. *Annu Rev Bioph Biom.* 25, 163–195 (1996).

444. H. Chen, S. Viel, F. Ziarelli, L. Peng, 19F NMR : a valuable tool for studying biological events. *Chem Soc Rev.* 42, 7971–7982 (2013).

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

446. C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl, G. J. Pielak, Protein 19F NMR in Escherichia coli. *J Am Chem Soc.* 132, 321–327 (2010).

447. W. Zhu, A. J. Guseman, F. Bhinderwala, M. Lu, X. Su, A. M. Gronenborn, Visualizing Proteins in Mammalian Cells by 19F NMR Spectroscopy. *Angewandte Chemie Int Ed.* 61, e202201097 (2022).

448. D. Xu, S. P. Meisburger, N. Ando, Correlated Motions in Structural Biology. *Biochemistry-us.* 60, 2331–2340 (2021).

CURRICULUM VITAE

Mark Vincent Carreon dela Cerna Biochemistry and Molecular Genetics 505 S. Hancock St., Louisville, KY 40205 (for full CV, visit <u>mvcdelacerna.xyz</u>)

ACADEMIC POSITIONS

Aug 2022 – Present	Assistant Professor of Biochemistry Georgia Southern University, Savannah, GA
Jan 2017 – Jul 2022	Lecturer, Adjunct Professor, Chemistry Bellarmine University, Louisville, KY
Aug 2016 – Jul 2022	Graduate Research Fellow, Biochemistry University of Louisville, KY

PUBLICATIONS

- [5] **dela Cerna, MVC**. "Reflections on Multimodal Delivery of a Laboratory Course for Non-Science Majors and Opportunities for Improved Student Engagement," *In Press: Journal of College Science Teaching (2022)*
- [4] dela Cerna, MVC; Garcia, RMM; Guidote, AM; Claudio, GC. "Molecular Dynamics Simulations of lota-Carrageenan Hexamer as a Single and Double Helix," *Journal of the Kentucky Academy of Science* (2022), 82: 32-41
- [3] Rivas, DR*; dela Cerna, MVC*; Smith, CN; Simpathi, S; Patty, B; Lee, D; Blackburn, JS. "A Screen of FDA-Approved Drugs Identifies Inhibitors of Protein Tyrosine Phosphatase 4A3 (PTP4A3 or PRL3)," *Scientific Reports (2021) 11, 10302*
- [2] Tyo, K; Duan, J; Kollipara, P; dela Cerna MVC; Lee, D; Palmer, K; Steinbach-Rankins, J. "pH Responsive Delivery of Griffithsin from Electrospun Fibers," *European Journal of Pharmaceutics and Biopharmaceutics (2018) 138: 64-74*
- [1] Bikash, P; Lin, BC; dela Cerna MVC; Hsu, CP; Lin, CH. "Face-to-face Paking of 2,3,9,10-Tetrasubstituted Pentacene Derivatives Revealed Through a Solid State [4+4] Thermal Cycloaddition and Molecular Dynamics Simulation," *Journal of Organic Chemistry (2016) 81: 6223-6234*

EDUCATION

Aug 2016 – Jul 2022	PhD. Biochemistry and Molecular Genetics University of Louisville, Louisville, KY Dissertation: Applications of Nuclear Magnetic Resonance Spectroscopy from Drug Discovery to Protein Structure and Dynamics Mentors: Donghan Lee, PhD; T. Michael Sabo, PhD; Brian Clem, PhD
Aug 2016 – Apr 2019	MS. Biochemistry and Molecular Genetics University of Louisville, Louisville, KY Project: Applications of Nuclear Magnetic Resonance Spectroscopy from Drug Discovery to Protein Structure and Dynamics Mentors: Donghan Lee, PhD; Brian Clem, PhD
Aug 2014 – Apr 2016	Grad Studies. Chemical and Physical Biology Vanderbilt University, Nashville, TN Project: Conformational cycle of dual function ABC transporter Mentor: Hassane Mchaourab, PhD
Mar 2012 – May 2014	MS. Chemistry (Physical and Inorganic Chemistry) Ateneo de Manila University, Quezon City, PH Project: Molecular dynamics simulations of carrageenan hexamers Mentors: Armando Guidote, Jr, PhD; Gil Claudio, PhD
June 2008 – Mar 2012	BS. Chemistry, minor in Economics Ateneo de Manila University, Quezon City, PH Project: Molecular dynamics simulations of carrageenan building blocks Mentors: Armando Guidote, Jr, PhD; Gil Claudio, PhD

SELECTED AWARDS AND HONORS

University of Louisville Nomination, NCI F99/K00 Transition Award
University of Louisville Nomination, HHMI Gilliam Fellowship
Mentored Undergraduate Research and Creative Awards Grant
University of Louisville Graduate Student Research Grant
Experimental NMR Conference Travel Fellowship
AAAS/Science Program for Excellence in Science
UofL Integrated Programs in Biomedical Sciences Fellowship
NIH T32 Molecular Biophysics Training Grant @ Vanderbilt
Vanderbilt Institute of Chemical Biology Fellowship
Philippines Department of Science and Technology Scholarship

SELECTED CONFERENCE PARTICIPATION

- [6] **dela Cerna, MVC**; Sabo TM, "Solution Dynamics of Human Guanylate Kinase by NMR," 2021, Annual Meeting of the Kentucky Academy of Science, KY
- [5] Elmoustapha, SE; dela Cerna, MVC; Lee, D, "Identifying small molecules targeting the MDMX-p53 interaction in human cancers," 2020, University of Louisville Undergraduate Arts and Research Showcase (part of Mentored Research Grant)

- [4] dela Cerna, MVC; Rivas, D; Brinegar, A; Burlison, J; Blackburn, JS; Lee, D. "Drugging an Undruggable: Targeting Pro-Metastatic Phosphatase PTP4A3," 2019, 4th Gateway NMR Conference, Ann Arbor, MI
- [3] Brinegar, A; dela Cerna, MVC' Lee, D. "Functional and Biophysical Characterization of Cancer-Related PTP4A3 Mutations," 2019, 4th Gateway NMR Conference, Ann Arbor, MI
- [2] dela Cerna, MVC; Ban, D; Arumugam, S; Burlison, J; Lee, D. "Towards Understanding the Role of Protein Motion in Oncoproteins that are Drug Targets," 2018, 59th Experimental NMR Conference, Orlando, FL
- [1] dela Cerna, MVC; Lin, BC; Lin, CH; Hsu, CP. "PM3 Simulations and Molecular Mechanics of Alkane-Substituted Pentacene: Preliminary Study of Solid State Pentacene Dimerization," 2014, Taiwan International Graduate Program – International Internship Program Presentations, Taipei, TW

PRESENTATIONS AND INVITED TALKS

* Invited

- [3] "Maximizing the Druggability of the Human Genome," Georgia Southern University, Savannah, GA, March 2022*
- [2] "A Tour of Structural Biology from COVID and Beyond and the (In)accessibility of Scientific Research," Philippine Science Communication Festival, Philippines, December 2021*
- [1] "Solution Conformation of Iota-Carrageenan: A Molecular Dynamics Study," 29th Philippine Chemistry Congress, Philippines, April 2014

SERVICE TO SCIENCE

- Kentucky Academy of Science, Newsletter Editor/Board Member (2020 Present)
- Louisville Regional Science and Engineering Fair Judge (2017, 2019, 2020)

SERVICE TO UNIVERSITY/DEPARTMENT

- Student Rep, UofL Medicine Faculty Forum (2021)
- **Student Rep**, UofL Biochemistry Graduate Education Committee (2019-2020)
- **President**, UofL Biochemistry GSA (2019-2020)
- Student Rep, UofL Graduate School Council (2018-2021)
- VP for Social Affairs, Vanderbilt Chemical and Physical Biology GSA (2015-2016)
- Organizing Committee, Vanderbilt Molecular Biophysics Seminar (2015-2016)

PROFESSIONAL MEMBERSHIP

- American Society for Biochemistry and Molecular Biology
- American Association for the Advancement of Science
- Biophysical Society
- Kentucky Academy of Science (ex officio board member)