12-13-2021

Utilizing Molecular Probes to Examine Protein Structure-Function Relationships

James Terrell

Recommended Citation

Terrell, James, "Utilizing Molecular Probes to Examine Protein Structure-Function Relationships." Dissertation, Georgia State University, 2021.

doi: https://doi.org/10.57709/25723719

Utilizing Molecular Probes to Examine Protein Structure-Function Relationships

by

James Ross Terrell

Under the Direction of Ming Luo, PhD

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Within the cell, the majority of functional roles are carried out by polypeptide macromolecules commonly referred to as proteins. Over the past 100 years, various methodologies to examine the high-resolution structural characteristics of proteins have been developed and utilized, including x-ray crystallography, nuclear magnetic resonance spectroscopy, and cryogenic electron microscopy. The structural information obtained has yielded much insight into the mechanisms by which many of these proteins function; however, limitations to the nature of the structural data have provided challenges when attempting to elucidate protein structure-function relationships and dynamics.

Owing to its nature, x-ray crystallography provides what can be described as a “snapshot” of the most favorable protein conformation or conformations in a particular crystal, which is affected by both the components of the solution as well as the crystallographic contacts that are necessary for the formation of the lattice. In this work, three different proteins have been examined crystallographically and their structures perturbed by the addition of molecular probes ranging in size from the miniscule molecular oxygen (hemoglobin), through the small molecule range of 50-1000 daltons (MDM2 RING domain), and well into the macromolecular range of >1 kilodalton (PU.1 transcription factor DNA-binding domain). Variances observed in the determined x-ray structures yielded valuable insight into the mechanisms by which these proteins function, as well as a deeper understanding of the methods by which they may be targeted for future treatment of human disease.

INDEX WORDS: Protein Structure-Function Relationships, X-ray Crystallography, Molecular Probes, Crystallographic Probes, Structure Perturbation

Utilizing Crystallographic Probes to Examine Protein Structure-Function Relationships

by

James Ross Terrell

Committee Chair: Ming Luo

Committee: Markus W. Germann

W. David Wilson

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

December 2021

DEDICATION

I would like to dedicate this work to my mother and father, Mary and Larry Terrell, and my wife, Kelly, for their unwavering love and support during my academic journey. Without you none of this would have been possible.

ACKNOWLEDGEMENTS

I would like to begin by acknowledging my academic advisor Dr. Ming Luo for helping to shape me into the scientist I have become. I began my doctoral career with an untamed passion for science that often sent me on tangential quests to find the next exciting thing (often before I had finished my original task). Although I am still guilty of this at times, your intelligence, pragmatism, and organized approach towards science inspired me and have given me the ability to approach scientific questions in a way which can lead to the exciting results I always seek.

I would also like to thank numerous members of the faculty and staff of the Georgia State University Department of Chemistry for their aid in all of my research efforts beginning with my committee members Dr. Wilson and Dr. Germann. Both of you have always been there for me when I needed to come and ask questions or to seek a new perspective on approaching research problems and I will be forever grateful. I would also like to give a huge thank you to our collaborators in Dr. Poon's lab and Dr. Poon himself for all of your help with the PU.1 project. Thank you to Dr. Jenny Yang, Dr. Binghe Wang and everyone else from the Center for Diagnostics and Therapeutics for their financial support and academic guidance.

I would also like to acknowledge the many friends and scientific colleagues I have had the fortune to interact with over the years. Thank you to my former lab mates Ian Davis, Kednerlin Dornevil, and Jiafeng Geng for helping to teach me about biochemical lab techniques and for making the Liu lab way more fun. Thank you to my current and former Luo lab peers, Scarlet Tang, Andrew Browne, Tales Rocha de Moura, Weiki Li, Chelsea Severin, and especially Ryan Gumpper. I enjoyed all of our wonderful conversations about science and life as we toiled away in the lab. It definitely made life more bearable through all the hard times.

To all the friends who have helped me along the way: Gala Chapman, Oswaldo Armas, Scott Dempsey, and all of the Georgia State Paintball folks… You have been more helpful than you can imagine. Keeping me sane even when things wouldn’t work in lab and just being there for me.

Most importantly I would like to thank my parents Mary and Larry Terrell, and my wife Kelly for everything they have done for me through my graduate journey. You have always offered me support and love and motivated me to keep going when I became discouraged. I hope I have made you proud and I will do my best to do great things with the knowledge I have gained.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Messenger ribonucleic acid (mRNA)

Transfer ribonucleic acid (tRNA)

N-formylmethionine (fMet)

Elongation factor (EF)

Release factor (RF)

Murine double minute 2 (MDM2)

Really Interesting New Gene (RING)

E26 transformation specific (ETS)

Winged helix-turn-helix (wHTH)

X-ray crystallography (XRC)

Nuclear Magnetic Resonance (NMR)

Cryogenic electron microscopy (cryo-EM)

M2 transmembrane domain (M2TM)

Cross-reactive material 197 (CRM197)

Human hemoglobin (HbA)

Molecular oxygen (O2)

Oxyhemoglobin (Oxy-HbA)

Deoxyhemoglobin (Deoxy-HbA)

Nuclear Factor of Activated T-Cells 1 (NFAT1)

Inulanolide A (InuA)

DNA-binding domain (DBD)

1 INTRODUCTION: STRUCTURAL AND FUNCTIONAL CHARACTERISTICS OF PROTEINS

When discussing biological systems, a common theme will often arise: That the structure of a macromolecule is closely tied to its function. From the very beginning of life on Earth, macromolecules have been essential in the processes that allow life to form, replicate and grow. Together through structure and function, molecular interaction, and the flow of energy; complexity beyond comprehension has arisen to produce incredible forms. At their core, all of these structures are commanded by the same basic chemical principles. By closely examining the structures of biological macromolecules we can begin to infer some of the information about the mechanisms by which they function. By understanding these structure-function relationships we can gain insight into complex methods for molecular synthesis, the underlying causes of diseases for which we can seek to develop therapies, potential vulnerabilities in detrimental pathogens, and even the mysteries of our origins as a species.

1.1 Protein Biosynthesis and Structural Features

1.1.1 Protein Translation

All organisms carry and replicate genomic material coding for proteins via 3-base codons which are transcribed by an RNA polymerase and subsequently translated by the ribosome [1]. The summary of all proteins produced by all reading frames and splicing products of a genome is known as the organismal proteome [2]. Proteomes can vary greatly in size depending upon the type of organism being examined. Viral proteomes represent some of the smallest proteomes, while multicellular organisms such as plants and vertebrates must produce many proteins to carry out all of the essential intra- and intercellular roles necessary for appropriate replication, metabolic balancing, cellular motility, and tissue formation [3]. The human proteome is currently known to code for over 20,000 individual proteins, but predictive models estimate that as many as 95,000 proteins may result following gene splicing [4]. The flow of information from an organismal genome to proteins is described in the central dogma of biology **[Figure 1.1]**.

Proteins are synthesized through the process of translation of a messenger ribonucleic acid (mRNA) template by the ribosome in cells. Following gene transcription, and appropriate pre-mRNA preparation in eukaryotes, protein synthesis occurs at the ribosome through the three stages of initiation, elongation, and termination [5]. Initiation of protein translation occurs with the formation of the initiation complex between the small ribosomal subunit, a captured mRNA containing an appropriate 5' sequence motif, such as the Shine-Delgarno sequence seen in prokaryotes or 5'-GTP cap in eukaryotes [6,7], and an initiator transfer ribonucleic acid (tRNA) carrying acyl-conjugated methionine in eukaryotes and N-formylmethionine (fMet) in bacteria, mitochondria and plastids [8,9,10]. With the formation of the appropriate initiation complex, large ribosomal subunit interactions envelop this ternary structure forming the exit (E), peptidyl (P), and aminoacyl (A) sites of the active ribosome [11]. Elongation begins when an additional tRNA containing the appropriate anti-codon recognizing the subsequent 3' codon enters the A-site and peptidyl coupling results from nucleophilic attack by the A-site tRNA amino group on the adjacent acyl group of the P-site tRNA. Breakdown of the resulting tetrahedral intermediate yields a newly formed amide bond [12]. The decoupled P-site tRNA and peptide bound A-site tRNA are then translocated to the E-site and P-site, respectively, aided by conformational changes induced by elongation factors (EF) through GTP hydrolysis [13]. This elongation and shifting process is repeated until a stop codon occurs in the mRNA strand which can be recognized by a release factor (RF) signaling for the hydrolysis and release of the newly synthesized peptide chain containing a terminal carboxyl group [14].

1.1.1 Primary Structure

The most basic structural element of a protein is known as the primary structure and represents the unbranched linearized amino acid sequence of the protein connected through amide bonds. The primary structure is comprised of 22 different L-amino acids in ribosomally synthesized polypeptides **[Figure 1.2, 1.4A]** [15]. Most protein structures incorporate 20 common amino acids, but unique tRNAs which recognize the *opal* (UGA) and *amber* (UAG) stop codons are used in rare instances to code for selenocysteine and pyrrolyine respectively [16,17]. Proteins synthesized by the ribosome through translation range in polymeric length from 20 to over 34,000 amino acids [18]. Structural elements at the primary level are derived from idealized planar conformations of the amide bond which is yielded by the delocalization of carbonyl electrons generating resonant sp^2 hybridization properties [19]. Torsion angles of covalent bonds from nitrogen_n (N_n) to the alpha carbon (Cα), Cα to carbon_n (C_n) position, and C_n to $N_{(n+1)}$ are represented by phi(φ), psi(ψ), and omega (ω) notations. Distribution plotting of backbone torsion angles φ and ψ yields energetically favorable "islands" which can be observed in the Phi-Psi plot developed by Ramachandran **[Figure 1.3A]** [20].

1.1.2 Secondary Structure

Secondary structural elements occur within proteins with replicative regions of favored backbone torsion combined with intrastrand hydrogen bonding between amide groups of the backbone **(Figure 1.4B)** [20]. Early fiber diffraction experiments by Astbury showed characteristics of helical properties in keratin fibers of wool and were denoted as the alpha (α) form of protein structure. Upon stretching, these α form fibers displayed an alternate diffraction pattern which was denoted as the beta (β) form [21]. In 1951, prior to any observable structure determination, a theory of potential secondary folding patterns was put forward by Pauling, Corey, and Branson using Pauling's theories of amide bond properties in their *PNAS* paper describing hydrogen bonding patterns of the classical α helix [22]. Soon after this landmark paper, Pauling *et. al.* proposed additional models describing the β-strand and β-sheet [23]. To date, theoretical works and structures observed in determined protein structures have allowed for the creation of a system describing observable secondary ordering; the Dictionary of Protein Secondary Structure (DSSP). The DSSP is often used when denoting secondary elements and is broken down into 8 categories describing local hydrogen bonding patterns and structural features computationally characterized and/or theorized **[Figure 1.3B]** [24]. α-helices and β-strands (parallel and anti-parallel) comprise the majority of secondary structural folding observed in proteins [25], but several other structures can result such as the 3_{10} helix, polyproline-II helix, π-helix as well as folds, turns, and loops which are essential in the formation of many functional motifs [26].

1.1.3 Tertiary Structures, Functional Motifs, and Domains

Combinatorial interactions between Cα-branching sidechains with alternate sidechains and regions of exposed amide backbone, combined with secondary elements of a protein structure and solvent interactions, give rise to increased ordering in the form of tertiary structures. Sidechain interactions promoted by hydrogen bonding, salt-bridges between ionized groups, van der Waals forces, and hydrophobic interactions can result in ordered folds known as structural motifs **[Figure 1.3C]** [27]. Additional stability can be aided by disulfide formation, cofactor binding creating apo and holo forms of motifs, and post-translational modifications [28]. Internal chemical bonding results in a protein displaying conformations in minimized energy states **[Figure 1.5A]** [29]. Energy wells derived from sidechain interactions resulting in altered tertiary ordering can lead to "misfolding" and result in a loss of biological function **[Figure 1.5B]** [30].

Full-length proteins may contain no tertiary elements or may contain multiple domains which bear tertiary structures linked by flexible regions. Protein domains contained within a single polypeptide chain may grant the protein multiple biological functions, act in tandem through individual roles to achieve a more complex biological function or act as autoregulatory elements which modulate the normal function of the protein when influenced by external factors [31]. An example of this is shown in the murine double minute 2 (MDM2) protein sequence **[Figure 1.6]**. The N-terminal region of MDM2 is known as the p53 binding domain and functions to both recruit p53 for ubiquitination by the C-terminal RING domain as well as bind and occupy the transactivation domain of p53 within the nucleus inhibiting transcriptional activity [32]. The C-terminal region known as the RING domain grants the protein E3 ubiquitination activity by recruitment of E2 ligases which can act upon bound p53 [33].

1.1.4 Quaternary Structures, Protein Complexes and Macromolecular Assemblies

Quaternary structural elements of proteins form upon oligomerization and interfacing of multiple protein chains. Formation of quaternary structure is observed in many biological systems and facilitates higher orders of protein cooperativity which regulate and mediate biological function [34]. An example of this is demonstrated in **Sections 2 and 3** describing the cooperativity of the α2β2 quaternary structure of hemoglobin as well as dimer formation in the MDM2 RING domain. Comparison of the structures of myoglobin and hemoglobin **[Figure 1.7]** demonstrates that both proteins contain nearly identical structures which display a classical globin fold, but quaternary assembly derived from oligomerization in hemoglobin yields an additional mechanism of affinity modulation allowing hemoglobin to more effectively disseminate oxygen from the lungs to tissues by allosteric control over ligand affinity [35].

Interfacing of multiple proteins with biomolecules such as DNA and RNA yields even larger complexes which can carry out roles vital to the persistence of the cell. **Section 4** describes the structure of the DNA-binding domain of the ETS-family transcription factor PU.1 which is utilized for the recruitment of cofactors and ultimately the formation of a macromolecular assembly with RNA polymerase which synthesizes an mRNA template [36].

1.2 Structure-Function Relationships in Examined Proteins

Biological functions of proteins range from simple structural roles such as scaffolding and stabilization all the way to greater orders of complexity such as catalysis, cellular transport, motility and signaling [37]. Closer examination of the proteins from **sections 2-4** display unique structural features described in **section 1.1** giving rise to protein functions.

Section 2 examines the crystal structure of hemoglobin. Functionally hemoglobin is utilized to bind and transport oxygen and the structure itself yields this ability via the incorporation of a heme cofactor as well as some more subtle characteristics which grant the ability to control O_2 affinity [38,39]. The globin fold utilized by hemoglobin to bind the heme co-factor is ubiquitous in nature and is observed in both prokaryotic and eukaryotic organisms [40]. The globin fold is comprised of 8 α-helices with flexible loops and turns between granting it a tertiary structure which localizes a hydrophobic pocket cradling the heme porphyrin with carboxyl sidechains extended towards the solvent exposed exterior. Proximal histidine coordination of the heme iron on the opposite face of the O_2 binding site forms a covalent coordination bond holding the heme cofactor in place and modulating O_2 affinity [41]. Cooperativity for oxygen binding and oxygen release takes place between the 4 subunits at the quaternary level through conformation changes induced by the ligand binding of O_2 and CO_2 as well as other allosteric modulators such as 2,3-

bisphosphoglycerate [42]. Together these structural characteristics grant hemoglobin the ability to bind and transport oxygen and deliver it to tissues.

The MDM2 RING domain examined in **Section 3** is a subdomain which acts as a recruiter for E2-ubiquitin conjugates [43]. The RING domain is observed in many other proteins which also function as E3 ubiquitination ligase mediators [44]. Structural examination of the secondary features demonstrate 3 short α-helices and 3 anti-parallel β-strands connected by flexible loop regions. The flexible loops between these structures contains two zinc-finger metal coordination sites comprised a combination of cysteine and histidine sidechains which aid in structural stability as evidenced by loss of function following metal chelation [45]. Dimer formation at the quaternary structural level is essential for E2 recruitment and has been a target for the design of drugs aiming to disrupt the hydrophobic dimer interface [46]. Recruitment of E2 ligases by this domain coupled with the binding of p53 to the N-terminus of the full-length protein facilitates p53 ubiquitination, ultimately granting MDM2 an essential regulatory role in p53 activity modulation which is observed with the death of mice containing *mdm2* gene knockouts [47].

The PU.1 DNA Binding domain from **Section 4** demonstrates an E26 transformation-specific (ETS) folding pattern. This structure displays a winged helix-turn-helix (wHTH) tertiary structure facilitating DNA binding by the insertion of an α-helix into the major groove of the DNA strand with sequence specific recognition granted by direct sidechain interactions with the consensus DNA bases [48]. Large clusters of lysine and arginine residues oriented towards the DNA phosphate backbone aid in increasing the overall affinity to the DNA. Structurally this motif grants the ability to bind DNA tightly at a desired site and coupled with actions of additional residues further into the N-terminal region of the protein, recruits protein co-factors which comprise a complex structure signaling the activation or repression of gene transcription [49].

1.3 Protein Structure Determination by X-ray Crystallography

Structure determination has been an essential tool allowing biologists and biochemists to understand the nature of the underlying chemical and structural interactions which give rise to macromolecular intra-, inter-, and extracellular functional roles. The combined fields of X-ray crystallography (XRC), nuclear magnetic resonance (NMR) spectroscopy, and cryogenic electron microscopy (cryo-EM) have allowed scientists to examine macromolecular structures at a high resolution and comprise the majority of over 181,000 depositions to the RCSB protein data bank (PDB) as of August 2021 **[Table 1.1]** with neutron diffraction and solution scattering being utilized to solve most of the remaining ~300 [50]. Advances in the fields of X-ray crystallography following determination of the first protein structure in 1958 by Kendrew *et. al.* involving data processing and concurrent advances in molecular biology, protein expression and purification methods, and X-ray data processing have lead to an explosion in the number of determined X-ray diffraction structures deposited annually since its inception in 1971 **[Figure 1.8]** [51].

1.3.1 Origins of Crystallography and X-ray Diffraction

X-ray crystallography has provided the scientific community with a wealth of information about the nature of molecular structures. The theory underlying crystallography is built upon a mountain of work dating all the way back to the 17th century efforts of Steno which demonstrated that crystals comprised of the same material display identical angles between crystal faces [52]. Based on this property Haüy was able to demonstrate that these faces could be derived from replicative subunits arranged in variable lattice patterns [53]. Just over 50 years later Miller used this information to denote the h, k, and ℓ integer values describing the relationship of crystal faces still in use today [54]. Following the definition of Miller indices came the works of Hessel, Bravais, Fedorov, Schonflies, and Barlow ultimately yielding the exhaustive description of all crystal symmetries physically possible and identifying 219 space groups (230 if chirality is included) [55-59]. Mathematically these advancements laid the groundwork which established the nature of crystal structure and symmetry, however, no method existed at the time to physically observe the true structure of a crystal.

X-ray structure determination traces its origins to 1895 with the initial characterization of X-rays by Wilhelm Röntgen via cathode ray tubes [60]. However, it wasn't until 1912 that Friedrich, Laue, and Knipping demonstrated that X-rays could be utilized to generate diffraction patterns of copper sulfate crystals [61]. Following their experiment, Laue proposed that the diffraction pattern was a result of an interaction between certain X-ray wavelengths with the atoms in the crystal lattice. Upon reading about Laue's observation, the father and son team of William Henry Bragg and William Lawrence Bragg used Laue's techniques and determined that X-ray reflections could be observed at determinate angles by planes of atoms in a crystal. The reflections observed can be described by Bragg's Law **(section 1.3.3)** [62]. With their new discovery, William Lawrence Bragg was able to determine the first crystal structure of a zinc blend crystal and thus the field of X-ray crystallography was born [63]. Ultimately their work on X-ray diffraction and structure determination lead to their award of the Nobel prize in Physics in 1915. Owing to difficulties in data analysis with increasing molecular and crystallographic complexity it would not be until 1957 that the first protein structure of myoglobin was solved by Kendrew under the advisement of Max Perutz [64].

1.3.2 Protein Crystallization

Protein crystallization is achieved by generation of a thermodynamically favored state in a supersaturated solution of protein. Dissolution of protein in an ordered fashion can result in crystal formation as long as a delicate balance between systemic enthalpy and entropy is yielded by formation of a crystalline state resulting in a minimized Gibb's free energy [65].

$$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$$

In the simplest terms the sum of the Gibb's free energy in a crystallization experiment (ΔG_C) yields a decrease in enthalpy which outweighs the decrease in entropy from decreases in protein mobility in solution.

The solubility phase diagram **[Figure 1.9]** demonstrates a simple explanation of the mechanism of crystal formation displayed in a vapor diffusion experiment, but in principle all techniques generating crystals result from a population of protein being driven through a concentration range at which the solvation is no longer favorable [66]. Crystallization occurs when a large population of the protein in solution is driven into a supersaturated interface between favored solubility and non-specific precipitation, phase separation or protein degradation [66]. High salt concentration in the solution or the addition of chemical precipitants which decrease protein stability or solubility can drive the protein into these solubility states [67]. This can occur through disruption of the solvation shell of the protein or by granting the solution protein denaturing properties resulting in regional agglomeration [68]. Modern crystallographic screening has been streamlined by the development of robotic methods to increase the number of conditions which can be screened simultaneously granting high-throughput capabilities for conditional optimization. Kits containing a broad array of precipitants, buffer conditions, and ionic components yield the ability to rapidly discern favorable conditions which can be scaled to grow large diffraction capable crystals [69]. Crystallized protein displaying sufficient uniformity of the lattice will generate observable diffraction upon X-ray irradiation. Crystals are typically treated with a cryoprotectant solution and flash frozen in liquid nitrogen or a stream of cold nitrogen gas prior to X-ray data collection in an effort to minimize radiation damage resulting in a loss of X-ray diffraction [70].

1.3.3 X-Ray Diffraction, Data Processing, and Model Building

X-ray irradiation of a protein crystal allows for structure determination by analysis of the resulting diffraction patterns. X-rays pass through electron clouds of each atom in a crystal structure and interactions result in elastic scattering [71]. When the scattered waves resulting from interactions of symmetrical atoms at separation distance (d) interact constructively they are observed as reflections described by Bragg's Law **[Figure 1.10]**:

$$n\lambda = 2dsin\theta$$

The resulting diffraction images as a crystal is rotated in an X-ray beam generates frames containing reflections with positions and intensities determined by relative orientation of atoms in the planes of the lattice [62]. Modern computational methods allow for rapid crystallographic indexing, yielding space group determination and unit cell dimensions by examining the relative positions between reflections in the diffraction image and determining which point group best accounts for the observed reflection pattern [70]. Integration of reflection intensities followed by scaling, merging and phasing yields an electron density map which displays an averaging of regions of electron occupancy in a unit cell **[Figure 1.11]** [72,73]. A model is then built into the electron density and the model is refined through iterative cycles in an attempt to account for observed electron densities [73]. Model validation can be achieved by examining reliability factors R_{work} and R_{free}, which examine how well the model fits the sum of observed density across the entire map and how well the map predicts a small set of unrefined test data [74].

1.4 Molecular Probes in Crystallography – Methods and Recent Discoveries

Native structures themselves have provided insights into numerous biological processes, but many of the proteins in the PDB have been examined in even further detail by modification or perturbation of their structures. The perturbation of a protein crystal structure can be accomplished via utilization of probes which influence intra- and intermolecular chemical interactions [75]. The definition of a probe in this sense is any alteration to the condition through which a structure has been determined in which a change can be observed and related to function. Selection of a probe is typically preceded by some insight into the functional role of a protein either through sequence homology, biological data, or a known substrate or ligand. Alteration of a protein's sequence by mutation, solution state changes in pH, incorporation of a ligand or small molecule, or co-crystallization of a protein complex can demonstrate structural shifts and yield valuable insights into the nature of a protein's structure-function relationships [76]. Incorporation of these molecular probes can occur during crystallization, or in many cases, following crystal growth depending upon the size of the probes and crystal packing [76]. Molecular diffusion through a crystal remains possible due to the large water networks which traverse the lattice, with typical solvent contents ranging from 27 to 78% in biomolecular crystals [77]. A description of the solvent content of a crystal can be derived by use of the formula yielding Matthew's coefficient V_M (specific volume, $Å^3$/Da) where V is the volume of the unit cell, n is the number of asymmetric units (ASU), and M is equal to the sum of the molecular weights of ASU contents [78].

$$V_M = V/(n * M)$$

Most experimentally observed values of V_M fall between 1.62-3.53 Å ^{3}/Da [77]. Specific solvent content can then be obtained by a simple operator:

$$Solvent\ content = (1 - \frac{1.23}{V_M})$$

Exploitation of these large water networks has been utilized for the incorporation of many of the molecular probes which will be described in subsequent subsections. Occupancies observed for small molecules are related to the dissociation constant (K_d) of the ligand potentially limiting the breadth of molecular structures which can be utilized [76].

Large or extremely hydrophobic molecules, or additional proteins or nucleic acids which interface with the protein target must be incorporated via co-crystallization [76],

1.4.1 Structure-function Insights Through Primary Sequence Variation

Substitution of amino acids within a protein sequence can demonstrate variability in activity and function of proteins. Individual amino acids may be substituted through alteration of the protein coding sequence prior to translation or by chemical or biochemical modification following expression if the residue is accessible [79,80]. Alteration of the primary sequence of a protein coupled with biochemical information can serve as a probe into how the amino acid sidechain participates in protein structure and biological function. Point mutations within a protein can yield structural insights into catalytic mechanisms and ligand binding sites and also chemical bonding networks which affect protein interfacing and dynamics [81,82]. Additionally, mutations can be used to mimic a post-translational modification such as phosphorylation which is heavily utilized for the regulation of many proteins by kinases and phosphorylases [83]. Numerous examples of structures determined containing mutations to amino acids demonstrated to be of functional importance through biochemical experiments or observed in disease states have been crystallized and their structures utilized to demonstrate specific perturbations altering protein function [76].

An example of this method was demonstrated by the examination of the MDM2 RING domain containing a phosphomimetic point mutation of S429E by Magnussen *et. al.* which was utilized to determine our structure described in **Section 3**. Comparison of the phosphomimetic MDM2 RING S429E mutant with that of the wild-type bound to UbcH5B-Ub in a ternary complex demonstrated the mechanism underlying the functional role of S429 phosphorylation in cells. The biological role of the MDM2 protein is to downregulate p53 activation of gene transcription [84]. The MDM2 RING domain aids in carrying out this role by acting as an E3 ligase bridging p53 bound to the N-terminal of the protein with E2 ligases carrying a ubiquitin conjugate [85]. Regulation of MDM2 is essential as disruption of p53 activity can be detrimental to autophagy regulatory mechanisms which prevent cancer and protect the cell [86]. DNA-damage induced activation of kinases results in phosphorylation of numerous residues in MDM2 [87]. Phosphorylation at residue S429 demonstrated an increase in autoubiquitination activity which can act on MDM2 to signal its nuclear export or degradation [88]. Determined structures demonstrated that S429 lies near the interface of the ubiquitin K33. Phosphomimetic interaction with K33 demonstrated increased stability of the N-terminus of the ternary complex in the previously proposed “closed” conformation leading to increased autoubiquitination rates yielding information about the structure-function role of pS429 [89].

Site-directed mutagenesis was also used to solve a 40-year mystery surrounding the mechanistic perturbation in the E52G mutant diphtheria toxin “cross reactive material 197” (CRM197). In the early 70s the Pappenheimer lab at Harvard sought to develop a less toxic diphtheria vaccine and was examining the effects of nitroguanidine mutagenesis on the function of the diphtheria toxin expressed by corynephage β infected toxigenic strains of *Corynebacterium diphtheriae* [90]. Mechanistically the diphtheria toxin functions to disrupt RNA translation by conjugation of an ADP-ribosyl group to the host cell eukaryotic elongation factor 2 (eEF-2) protein which is essential for the elongation step of protein synthesis [91]. The group noted that among the many mutations induced by nitroguanidine that the 197th "Cross Reacting Material" characterized had lost nearly all toxicity which resulted from a glycine substitution of glutamate 52 (E52G) in the A toxin subunit [90]. Owing to its lack of toxicity but retention of toxoid uptake immunogenicity this "CRM197" was utilized in the 1990s by Wyeth to develop conjugate vaccines against other bacterial infections with the eventual marketing of the Prevnar® *Clostridium pneumoniae* vaccine [92]. However, it wasn't until 2012 with the work of Malito *et. al.* that the structural basis for this lack of activity was characterized crystallographically. Upon structural determination the novel E52G structure was compared with that of the wild-type diphtheria toxin. Their findings demonstrated that although the E52 residue of DT is not directly involved in NAD hydrolysis necessary for toxicity, that increased flexibility of the loop containing E52 significantly disrupted NAD binding leading to a drastic reduction in toxoid functionality [93].

1.4.2 Structures at Alternate pH

The chemical bonds through which protein structure is derived can be variable and pH dependent. Sidechains of aspartate, glutamate, histidine, lysine, arginine, cysteine, and tyrosine are all ionizable as pH rise and fall [94]. The local environment of the sidechain can alter pK_a values and canonically basic residues such as lysine can have pKa values drop as low as 5.3 in a unique protein folding environment [95]. Physiological pH inside human cells typically falls between 7.0 and 7.4, but some organelles derive their functions through pH variation outside of this range. Lysosomal and phagocyte pH are both much more acidic than the cytosol with lysosomes reaching pH values as low as 4.5 [96], while mitochondrial matrixes contain a pH of approximately 8.0 facilitating ATP synthesis through proton motive force [97]. Proteins which are localized to variable pH ranges throughout their lifespan may utilize the environmental pH to induce conformational changes for molecular activation or repression [98]. Altering the solution pH during crystallization can result in the growth of a crystal displaying alternative packing and variable structural characteristics altering functionality [99]. Structure determination of proteins across a broad pH range may be utilized to examine some of these structure-function relationships. An example of this effect has been demonstrated with the determination of pH-dependent conformational changes in influenza A M2 matrix protein which alter the interior water network [100].

A primary mechanism through which Influenza A virus enters the cell upon surface binding is by exploiting clathrin-mediated and alternative methods of endocytosis [101]. Acidification of the endosome activates conformational changes in viral proteins which mediate viral structure and membrane disruption leading to cytosolic release of virion contents [102]. Early acidification of the endosome activates conformational changes in the M2 matrix protein which is comprised of four 97 residue subunits that traverse the viral envelope [100]. A water network conserved in the core of the tetrameric transmembrane domain (M2TM) functions as a proton mediation network facilitating rapid H^+ transfer occurring via the Grotthuss mechanism [103]. With a drop in endosomal pH, successive protonation of 4 histidine residues (His37) in the tetrad which form a “His-box” [104] located at the opening of this pore accumulate a positive charge which can be transferred to a water network that becomes more compact and ordered at low pH [105]. Acidification of the interior of the virion promotes the dissociation process between viral ribonucleoprotein (vRNP) complexes and the M1 matrix protein [106]. This prepares viral contents for release following membrane fusion induced by pH-induced conformational changes in the hemagglutinin protein [107]. The blocking of this mechanism by amantadine was one of the earliest methods of antiviral treatment for influenza infection [108]. Recent news has highlighted the controversial use of hydroxychloroquine (HCQ) for the treatment of the SARS-CoV-2 virus which has been demonstrated to prevent acidification of the endosome and inhibiting viral replication *in vitro* [109]. Unfortunately, off-target effects and an insufficient therapeutic index have ruled out HCQ as a viable clinical treatment for the disease [110], but the observed mechanism may be useful for future therapeutic development.

1.4.3 Small molecules and ligands

Among proteins small molecule interactions and effects are widespread. Small molecule binding is facilitated through interactions with complementary components of the protein generating an energetically favorable ($-\Delta G$) interaction represented with the K_d dissociation equilibrium constant [76]. Small molecule binding by proteins is essential for cellular replication and energy production and the synthesis of complex small molecules in enzymatic pathways. Ligand binding is also utilized by proteins for regulatory and signaling roles [111-112]. Although suspended in a crystal lattice, proteins still possess the ability to undergo conformational changes and facilitate molecule binding unless it is prohibited by interfacial contacts between protein subunits [76]. Countless examples of ligand bound structures have been published and many enzymes have been demonstrated to retain catalytic activity *in crystallo* with observation of catalytic intermediates observable by freezing of crystals following molecular incorporation at varied time intervals [112]. Incorporation of small molecules either by soaking or co-crystallization can be used to examine ligand binding sites and allosteric structure-function relationships. Insights from these methods has granted the development of structure-based drug design methods by identifying drug binding sites allowing for analog synthesis with potentials for enhanced activity or reduction of off-target effects [113]. This technique was utilized in **section 3** to examine drug binding sites in the MDM2 RING domain structure.

One of the earliest examples of structural perturbation by ligand binding was conducted on the subject of **section 2**: Hemoglobin

Early studies on hemoglobin demonstrated that the protein possesses a sigmoidal O_2 binding curve. This suggested that cooperativity must occur between the four subunits in some way [114]. Perutz *et. al.* determined the structure of oxyhemoglobin in 1960, however, it would be several more years before the deoxy- structure would reveal the key differences between the relaxed (R) and tense (T) states observed with the O_2 ligand binding site occupied [115]. Deoxy structures displaying the T state undergo a conformational change induced through the proximal histidine opposite the O_2 binding site. This coordinate histidine (HF8) induces movement in the backbone of helix 8 which is transferred to the adjacent structure through the loop region at the end of the helix. This drives the protein into the R-state and opening the distal histidine (HisE7) gates which are the primary regulators of O_2 entry into the heme site. This cooperativity between subunits grants hemoglobin a unique characteristic for the transport and delivery of O_2 to other tissues where the partial pressure of O_2 is diminished and O_2 release promoted by allosteric effects of CO_2 binding at alternate sites [116]. Further studies using ligands would go to to demonstrate the important role of 2,3-bisphosphoglycerate as a regulator of O_2 release which structural determination revealing its binding site at the core of the tetramer between the four subunits [117].

1.4.4 Protein complexes

Protein interactions with additional proteins and other macromolecular structures can be examined by combining the two molecules in a crystallization droplet at an appropriate stoichiometric ratio granting complex formation and allowing co-crystallization [76]. Comparison of the bound and unbound states can yield valuable insight into the underlying mechanism of complex formation and the nature of macromolecular epitopes which yield biological functions [118]. Understanding of these macromolecular epitopes has been utilized to gain mechanistic information about immune functions, DNA and RNA binding, protein regulatory and transport mechanisms, and macromolecular assembly [76,118-120]. Utilization of this method has also employed antibodies for protein stabilization and granted the ability to examine structures which would otherwise be unable to be crystallized [121]. **Section 4** utilized complex co-crystallization to examine the structure of the DNA-binding domain of the PU.1 transcription factor bound to DNA strands.

A protein-protein ternary complex between the MDM2-MDMX RING heterodimer was published in 2017 granting the first structural revelations into the structure-function relationships which give MDM2 its ability to act as an E3 ligase by bridging the structures of p53 and a recruited E2-Ub conjugate[122]. In 2020 the same group determined the first homodimeric structure both bound and unbound to the recruited E2-Ub [89]. The epitope revealed subtle differences between the MDM2 homodimer interactions with UbcH5B-Ub and MDM2-MDMX heterodimer interactions and yielded insights into the variability between activity of these two complexes and the ability to target them with some of the drug molecules we examined in **section 3.** The extreme C-terminal interactions in the heterodimeric structure create more contacts with the ubiquitin carried by the E2 ligase which would potentially block the active site of our compounds[89,122]. This may explain the lack of activity by both InuA and MA242 against the heterodimer. Both structures demonstrated contact between the UbcH5B-Ub at both of the Zn-finger motifs also yielding insight into how zinc chelators eliminate the activity of MDM2 [123]. Both extreme N-terminal regions of the dimers extend outwards into the solvent where the flexible linkers further into the MDM2 sequence will connect with the acidic and p53 binding domains.

1.4.5 Temperature During Irradiation

Recent advancements in the field of crystallography have demonstrated that structure determination at room temperature can reveal previously unobservable protein conformations and dynamics [124]. Diffraction at cryogenic temperatures has become a common practice which is used to minimize the damage induced by irradiation of a crystal over time and has been implemented in the determination of more than 95% of all crystal structures [125]. However, time scales of crystal freezing relative to conformational changes in proteins occur on different orders of magnitude and the freezing of crystals may result in the protein subunits settling into energetically favorable states which are transient when a biological role is being fulfilled [126]. This lack of conformational variability contributing to protein function can be lost at cryogenic temperatures. Experiments examining diffraction of myoglobin have demonstrated this effect and a change in the volume of the protein and interactions between α-helices were noted between diffraction observed at cryogenic temperatures to those obtained at room temperature [127]. In their work Fraser *et. al.* examined diffraction data from 30 protein structures determined at cryogenic and room temperatures. The proteins cyclophilin A (CypA) and the HRas GTPase both demonstrated alternative conformations in sidechains involved in substrate binding and enzymatic activity demonstrating that the temperature at which data collection takes place may also serve as a valuable probe into sidechain dynamics and protein conformations owing to effects on the protein structure induced by crystal freezing [124].

1.5 Research Aims

The research contained in the following chapters utilized several of the above-described molecular probes to examine a broad range of protein targets involved in human disease states **[Figure 1.12]**. Crystallographic examination of the three protein systems incorporating molecules which perturbed their structures was achieved. Additional experiments employing biochemical and biophysical methodologies supported the observed structural effects and gave credence to the proposed structure-function relationships. Insights into the mechanisms underlying the structure-function roles in these proteins aimed to achieve:

1. A greater understanding of gas channels in human hemoglobin and the correlated protein mutations resulting in disease states
2. The identity of a structural site for disruption of the E3 ubiquitin ligase activity of the murine double minute 2 (MDM2) protein
3. An understanding of the mechanisms by which the transcription factor PU.1 recognizes DNA sequences and the effects of flanking sequences on affinity

Aim 1 was achieved following the development and implementation of a novel method for the crystallographic incorporation of large quantities of molecular oxygen via liquid oxygen (LO_2) crystal soaking. The mechanisms through which hemoglobin binds and transports O_2 have been described in great detail, however, the channels through which O_2 reaches the heme co-factors are not fully understood. Computational modelling of O_2 diffusion pathways as well as experiments incorporating elemental xenon into crystal structures have sought to provide some insights into gas diffusion channels, but no direct evidence of specific O_2 diffusion channels had been produced [128,129]. Structural perturbation of hemoglobin *in crystallo* by an influx of O_2 was observable in the form of crystallographic B-factor increases and regions in which increases were observed directly correlated with mutation sites associated with human disease states. This is the first example in the literature in which liquid oxygen has been utilized to perturb a crystal structure and provides crystallographers with a novel method to examine proteins utilizing molecular oxygen.

Aim 2 was accomplished by examining crystal structures of the MDM2 RING domain. The MDM2 RING domain acts as the E3 mediator between coupled E2 ligases and p53 during ubiquitin transfer. Ubiquitination of p53 is implicated in numerous cancers as it will both downregulate its transcriptional activity and will also signal its degradation. Several compounds binding the MDM2 RING domain have been demonstrated to alter the E3 activity of MDM2 towards p53, but there has been no structure presented to date which displays mechanistic insights into how these inhibitors alter protein function. Crystals incorporating two inhibitors were examined and both demonstrated a common binding site. The site at which both bound was near the extreme C-terminus which has been demonstrated to be essential in E3 activity. Comparison of our structures with ternary E2-Ub-MDM2 complexes demonstrated that Tyr489 interfacing is essential in ternary complex formation and biochemical experiments employing our inhibitors demonstrated disruption of ternary complex formation through a decrease in ubiquitin transfer rates. These experiments yielded insight into both the structure-function relationship of Tyr489 as well as mechanistic information on RING inhibition which can be used to develop better therapeutic agents.

Aim 3 utilized the PU.1 DNA-binding domain co-crystallized in complex with large probes in the form of high- and low-affinity DNA sequences. The determined structures demonstrated the highest resolution structures of an ETS transcription factor observed to date and the drastic improvement in resolution, as well as the perturbations in the structure observed with flanking sequence variation, gave insights into key residues involved in DNA recognition and affinity modulation. MD simulations further established the roles of residues displaying conformational variability between the two structures and a crystallization condition was developed through which many other sequences may be examined in the future.

Together these experiments demonstrated that molecular probes could be incorporated crystallographically into the systems being examined and yielded valuable insights into the structure-function relationships of proteins that were utilized.

1.6 Figures

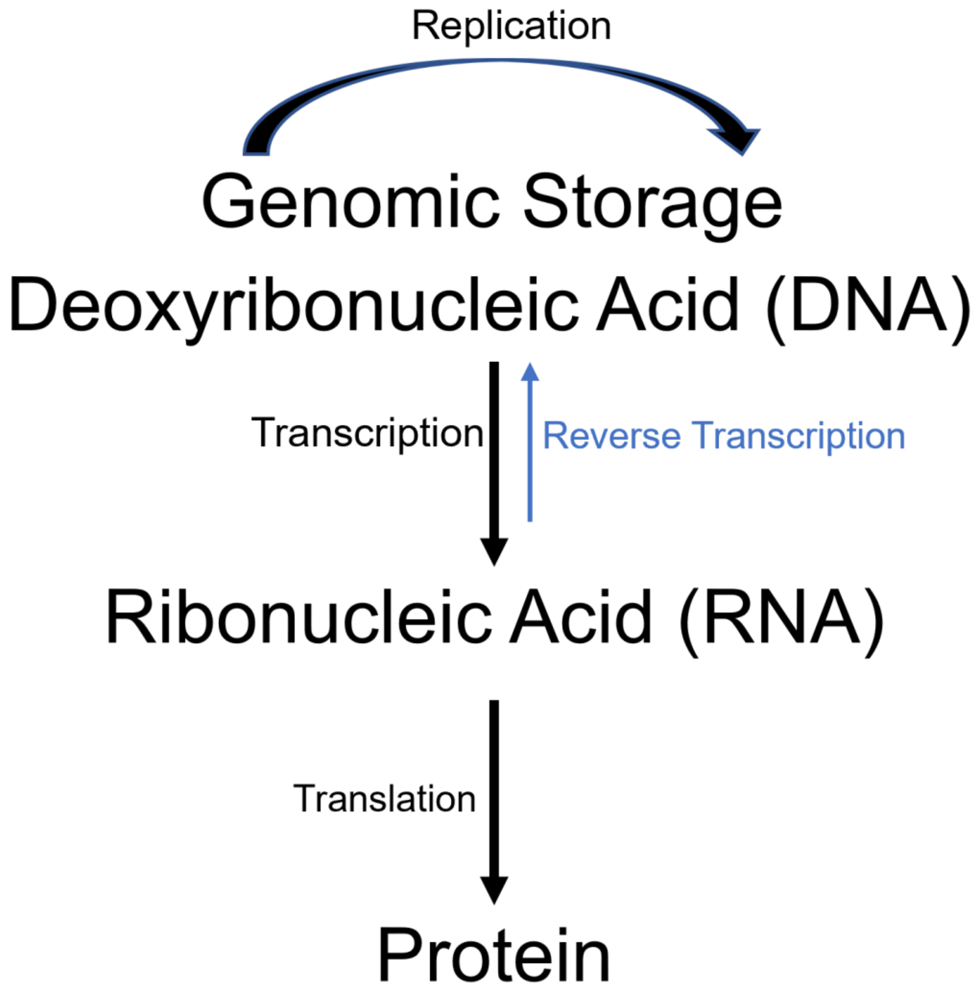


Figure 1.1 The Central Dogma of Biology
The central dogma of biology describing the flow of information from genomic storage (DNA or any number of combinations of positive and negative, single and double stranded, RNA or DNA/RNA hybrids) to an mRNA template to a protein following translation by the ribosome

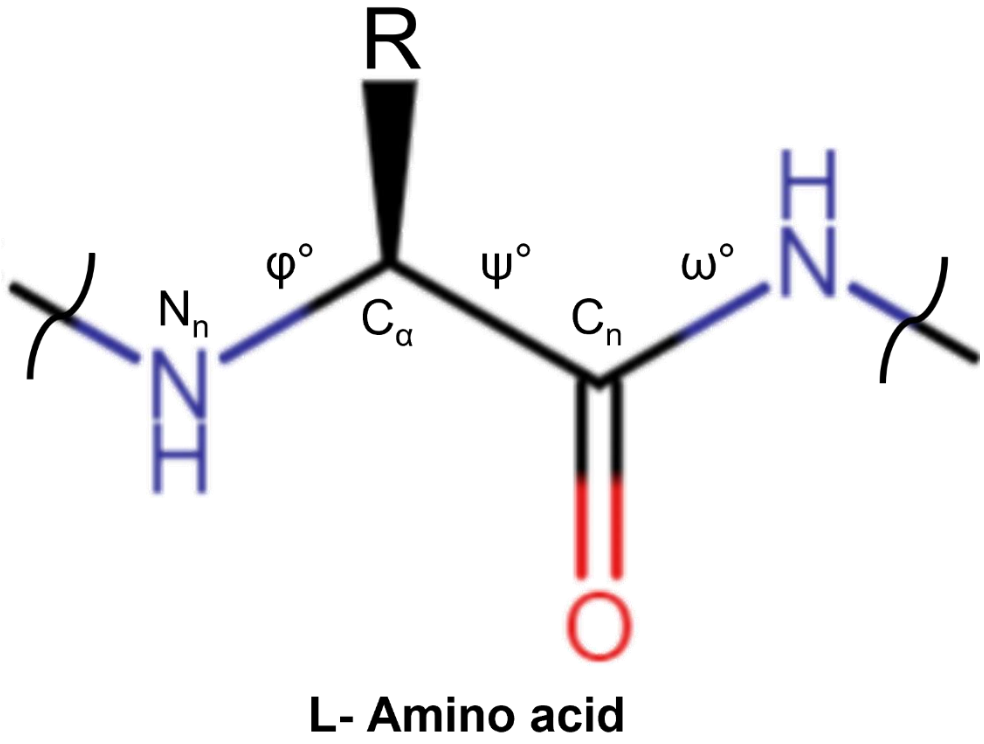


Figure 1.2 2D representation of an L-amino acid polymerized in a peptide chain
The N_n, alpha carbon (C_α), and C_n positions are denoted. Torsion angles φ,ψ, and ω are shown at their appropriate bond postions

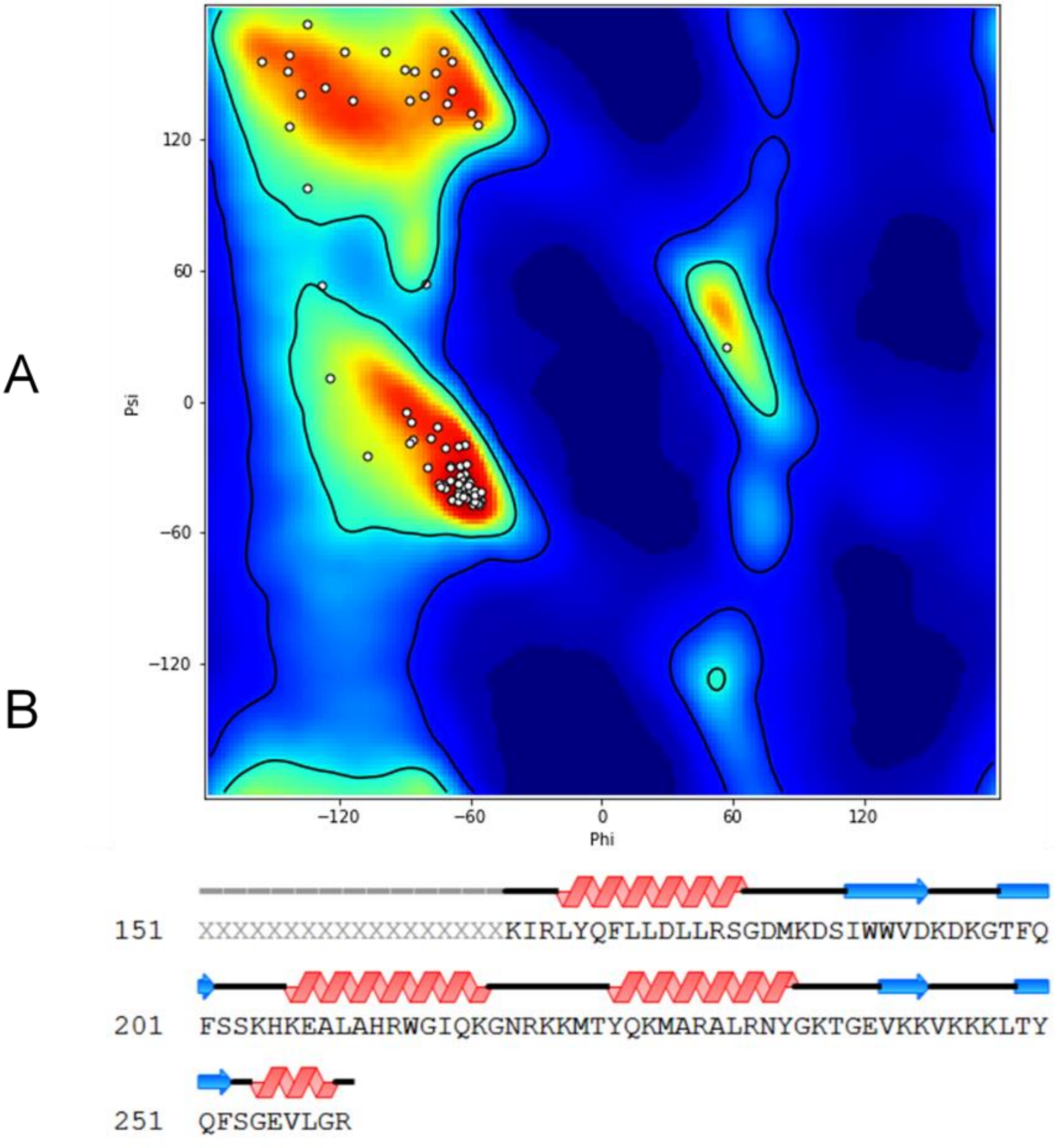
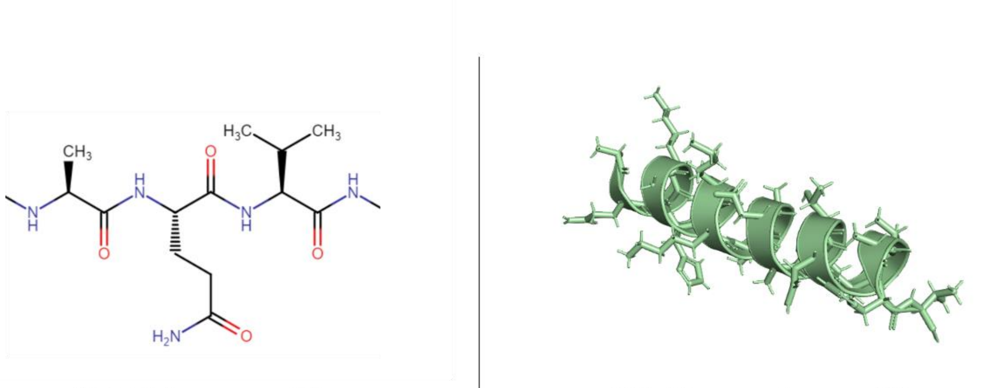


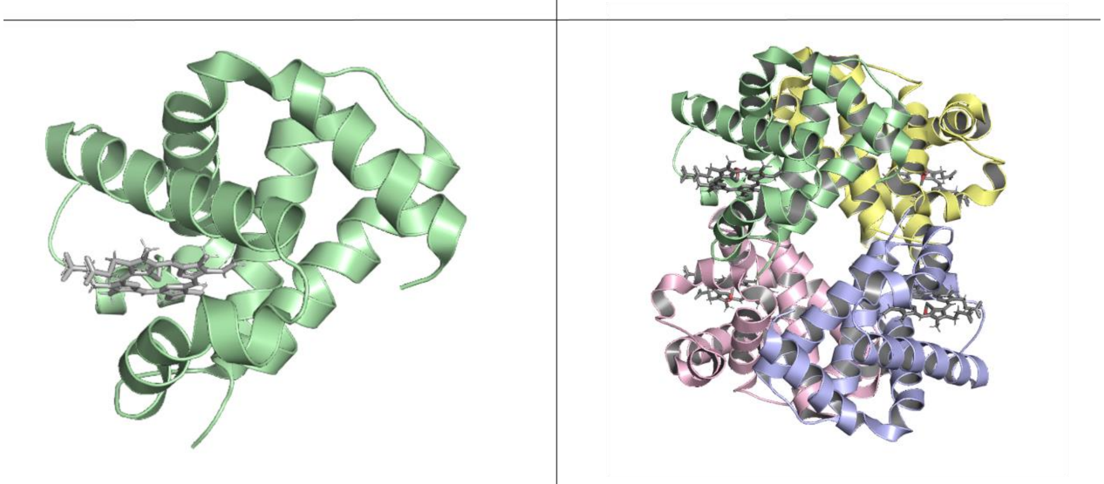
Figure 1.3 Ramachandran plot and secondary structure plot of the low-affinity structure described in section 4.3

A Ramachandran plot displaying the comparison of φ and ψ torsion angles (°) of each amino acid in the determined structure of PU.1 DNA-binding domain. **B** Secondary structural elements of the protein corresponding to their positions in the protein primary sequence. **A** and **B** were generated in the phenix.refine module of PHENIX software [73]



(A) Primary - 1º

(B) Secondary - 2º



(C) Tertiary - 3º

(D) Quaternary - 4º

Figure 1.4 Structural hierarchy of proteins
A The primary structure of a protein comprised of amide-linked L-amino acids **B** An α-helical secondary structural element of human hemoglobin. **C** The tertiary structure of an α subunit of human hemoglobin containing a heme co-factor and displaying the classical globin fold **D** Quaternary ordering observed from the oligomerization of four hemoglobin subunits displaying the α2β2 oxygen bound R-state (PDB ID: 6BB5)

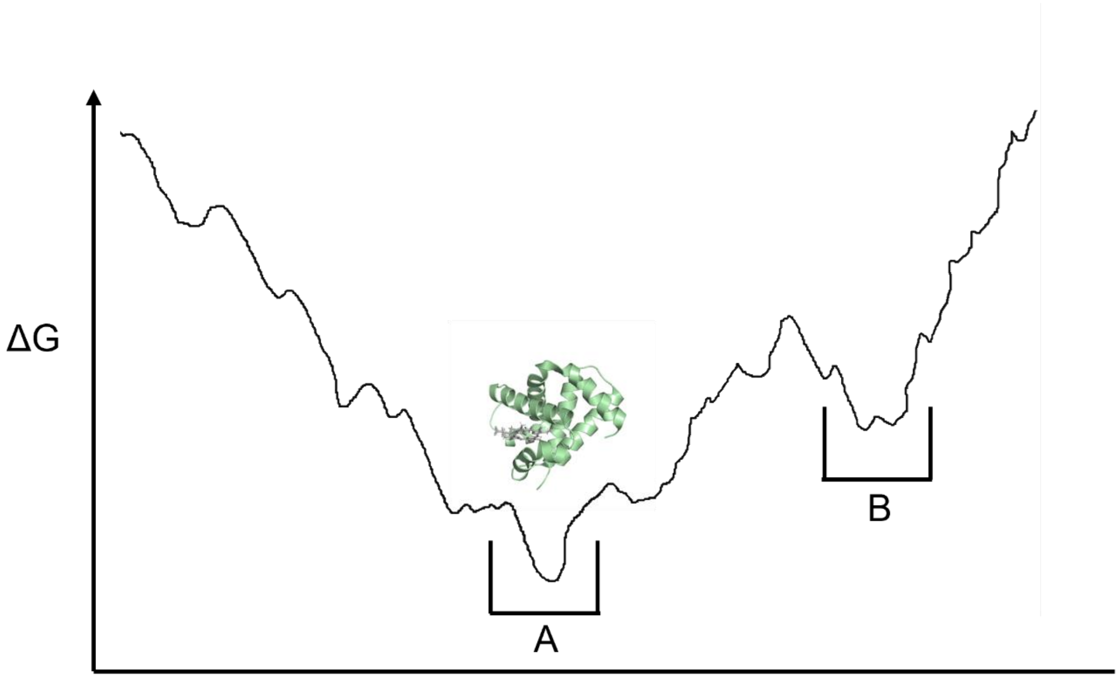


Figure 1.5 Energy funnel representation of protein folding states
A Free energy “wells” result in stabilized protein folds which will retain their structure unless the environment significantly disrupts internal chemical interactions. **B** Alternative folding states may result in a protein conformation falling into an energy well in which biological function is disrupted by lack of proper motif formation

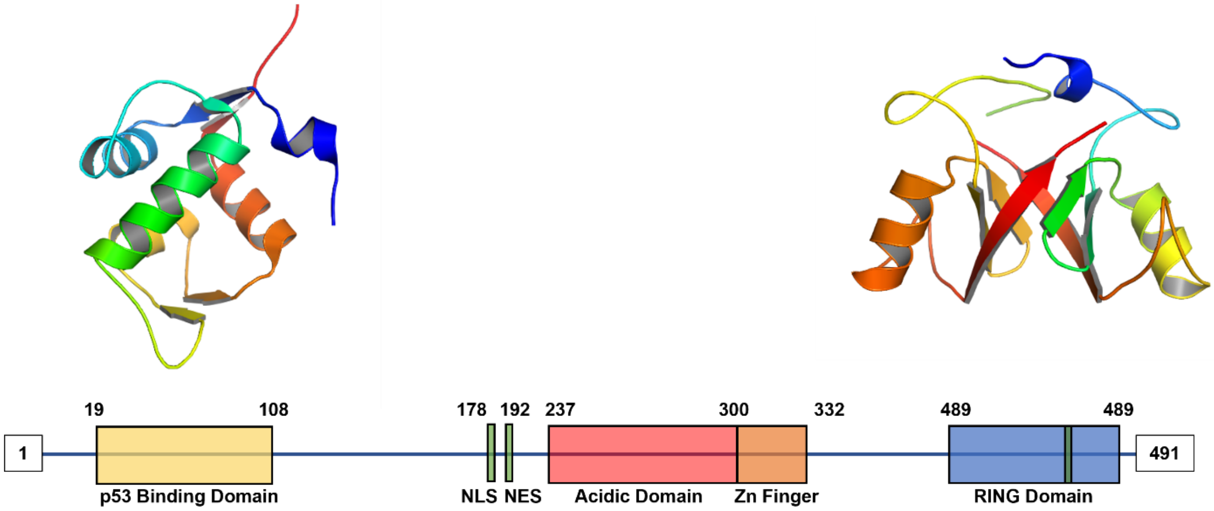


Figure 1.6 MDM2 protein domains and associated structures
The MDM2 protein contains 491 amino acids with 4 protein domains connected by flexible regions. The N-terminus binds p53 reducing its transcriptional activity while the C-terminal RING domain recruits E2 ligases for the ubiquitination of bound p53. (PDB ID: 1YCR)

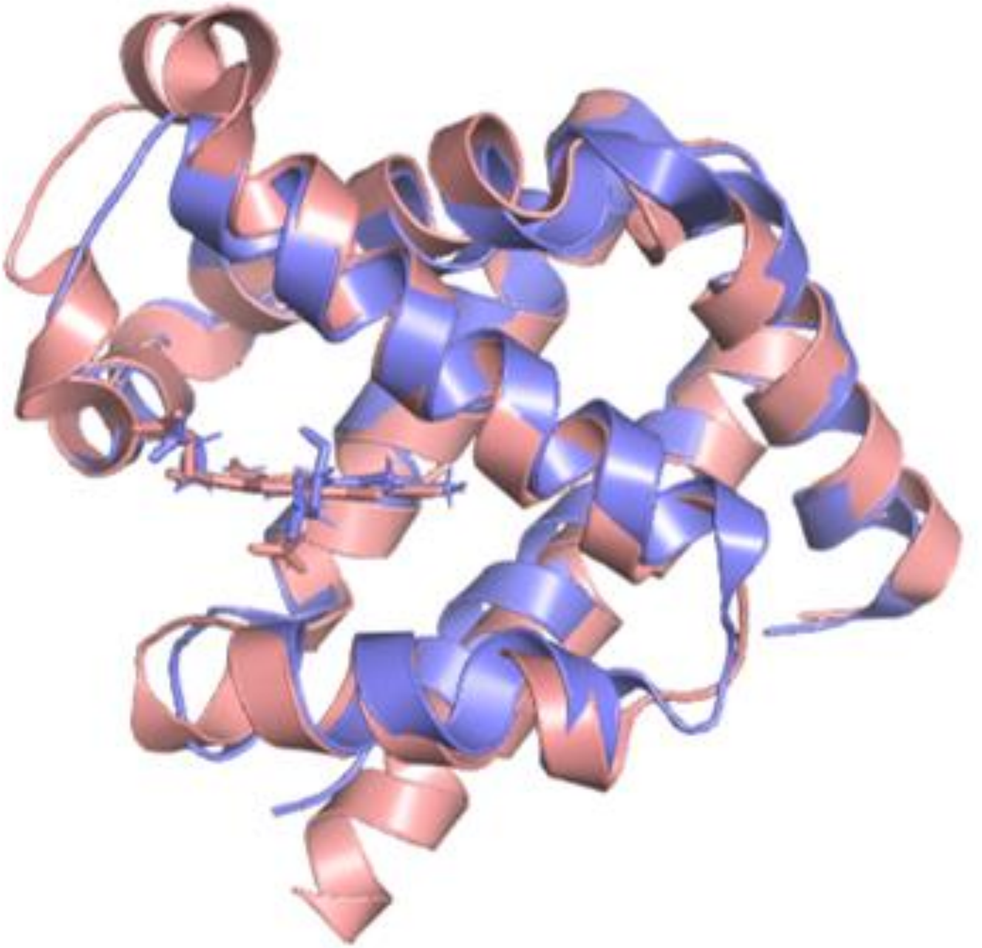


Figure 1.7 Structural overlay of myoglobin and the hemoglobin α subunit
Myoglobin (salmon, PDB ID: 3RGK) and Hemoglobin (Slate, PDB ID: 6BB5) both contain a globin fold shared by many members of the globin superfamily. Structural variance in hemoglobin allows for tetrameric quaternary structure formation which allows for cooperative O_2 binding between the oligomeric units

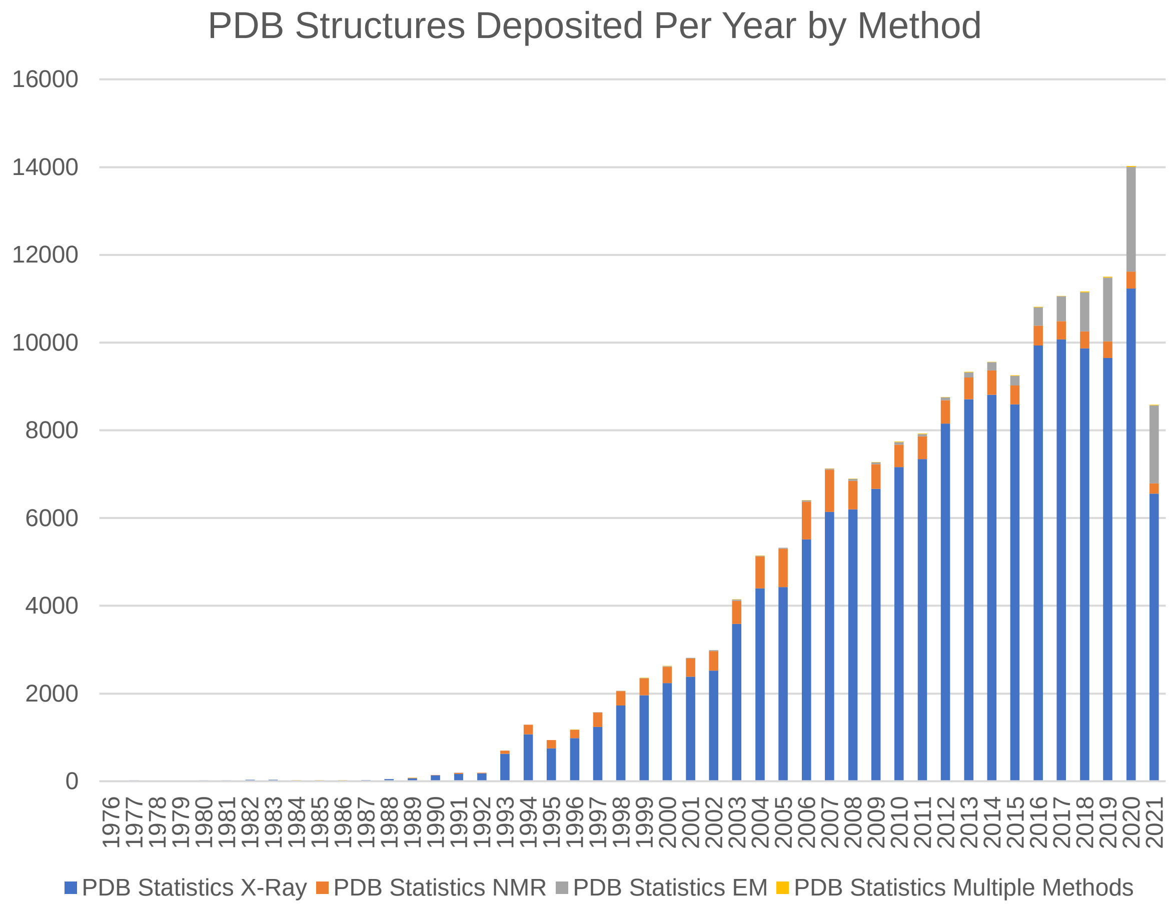


Figure 1.8 Structural deposits to the RCSB Protein Data Bank by year and method
Annual deposits to the protein databank have grown rapidly since its inception in 1971. Over 14000 structures were deposited in 2020 alone. Increases in accessibility to cryoEM as well as advances in structure determination have led to a rapid growth in cryoEM deposits over the past 10 years.

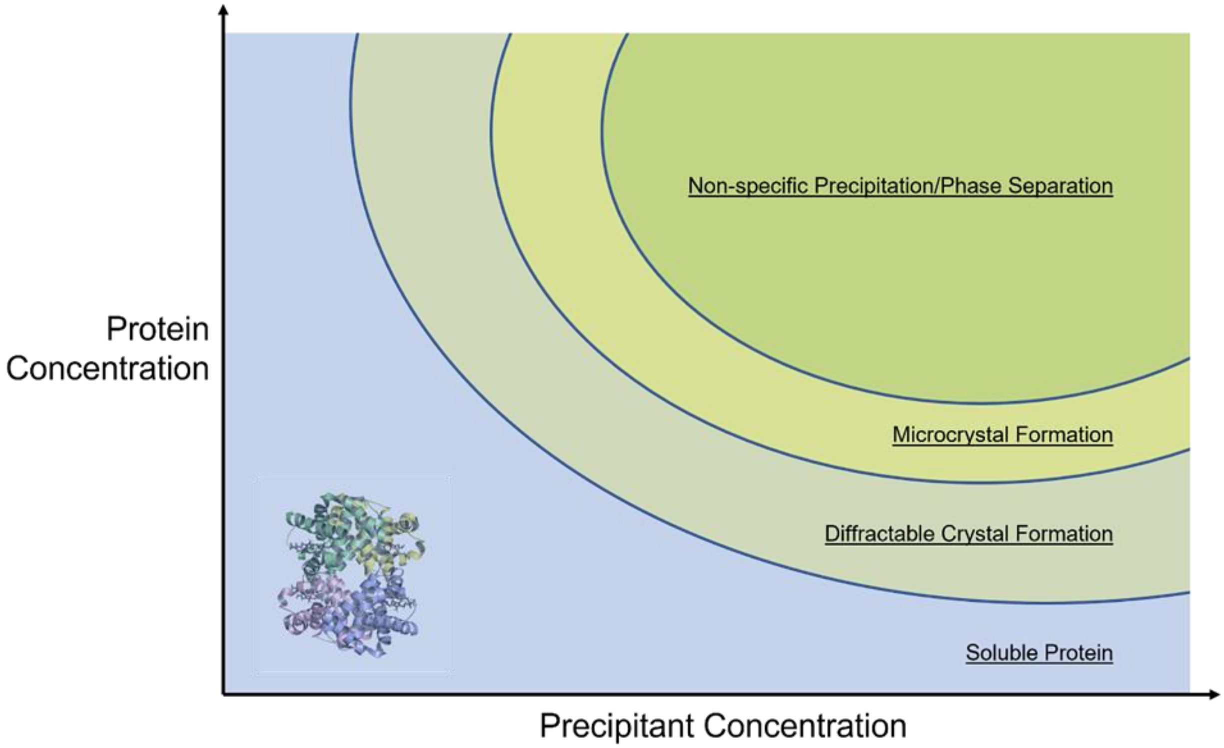


Figure 1.9 Solubility phase diagram during a vapor diffusion experiment
Protein contained in the solution is driven into metastable and supersaturated zones by increased protein and precipitant concentration over time as water diffuses from the crystallization droplet. Non-specific precipitation and/or phase separation often result if conditions are not optimized. Ideal conditions maintain the supersaturated/metastable state which over time will drive protein to crystallize

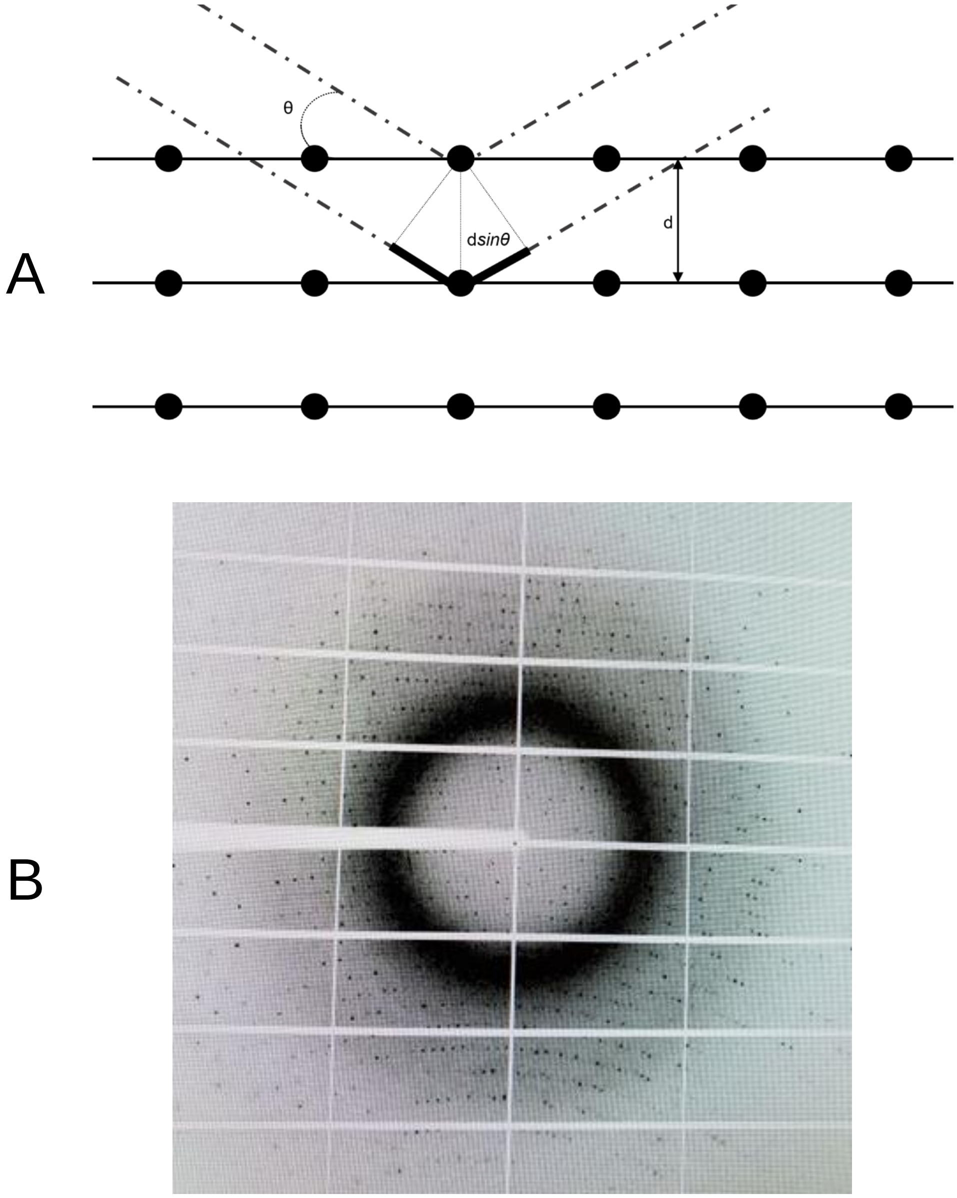


Figure 1.10 Bragg diffraction diagram and diffraction resulting from X-ray diffraction by a PU.1 crystal
A The principal of Bragg diffraction displaying the direction of constructive interference between scattered X-rays. **B** Observed reflections resulting from irradiation of a PU.1:DNA complex crystal on a Dectris EIGER X 16M detector.

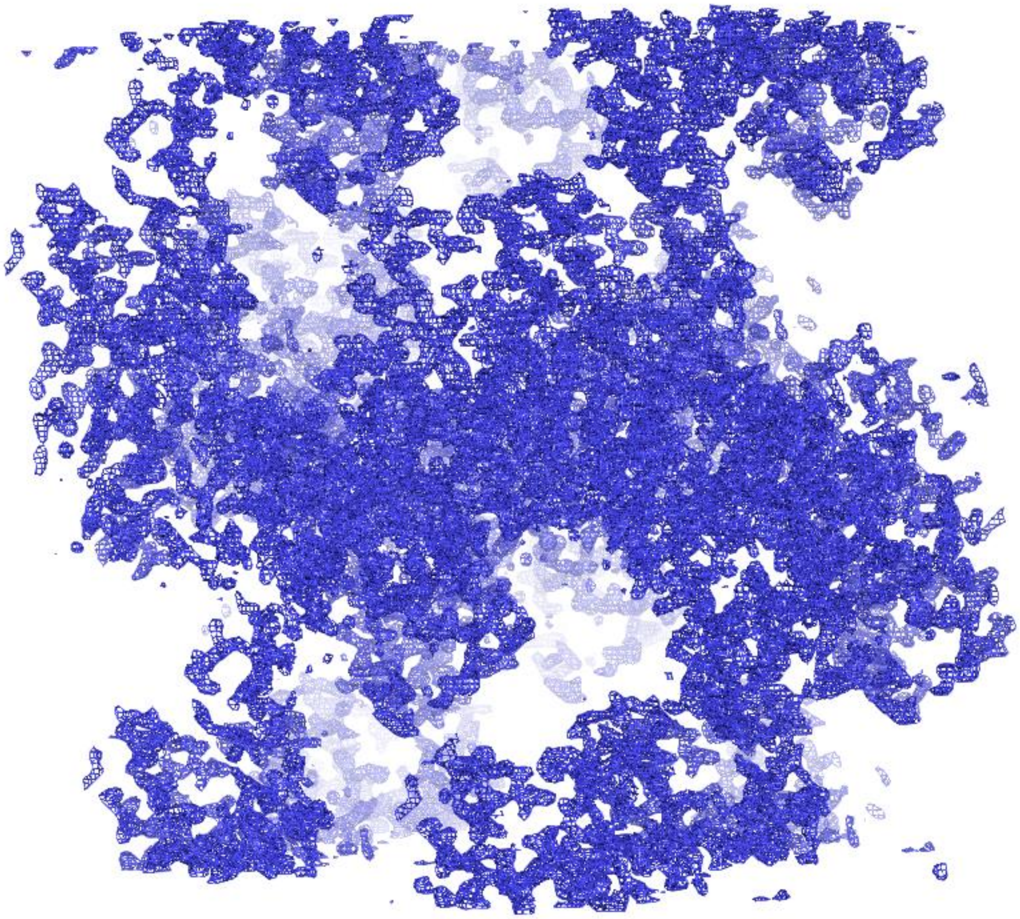


Figure 1.11 Electron difference map without a model input
After data processing and phasing from single anomalous dispersion or molecular replacement, an electron density map can be constructed like the one depicted above. Models are built into the density and refinement generates statistics for validation purposes

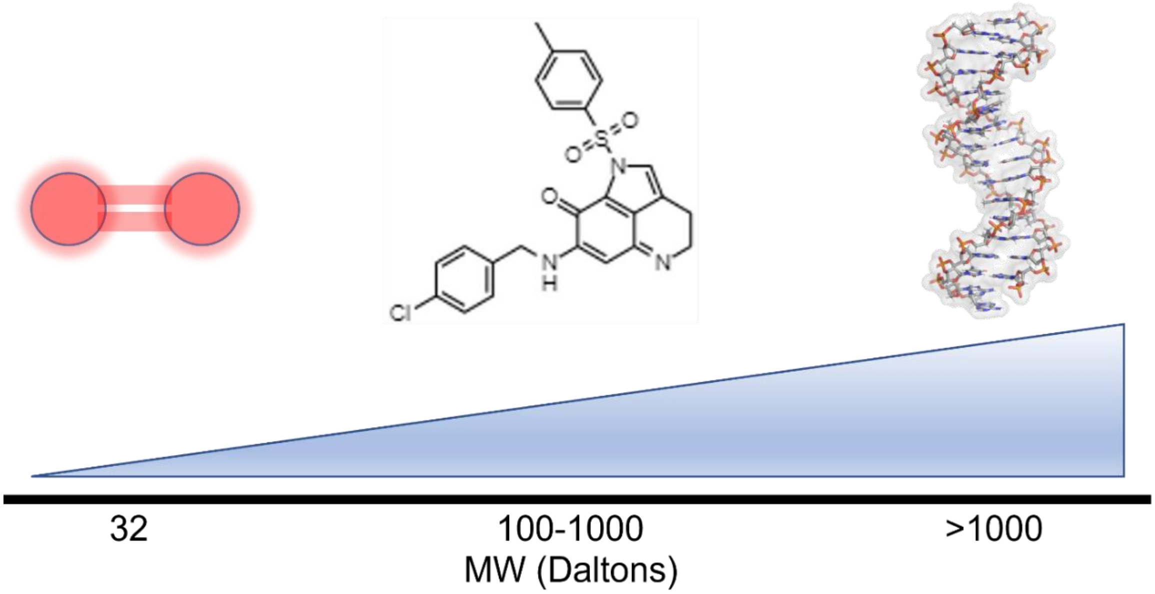
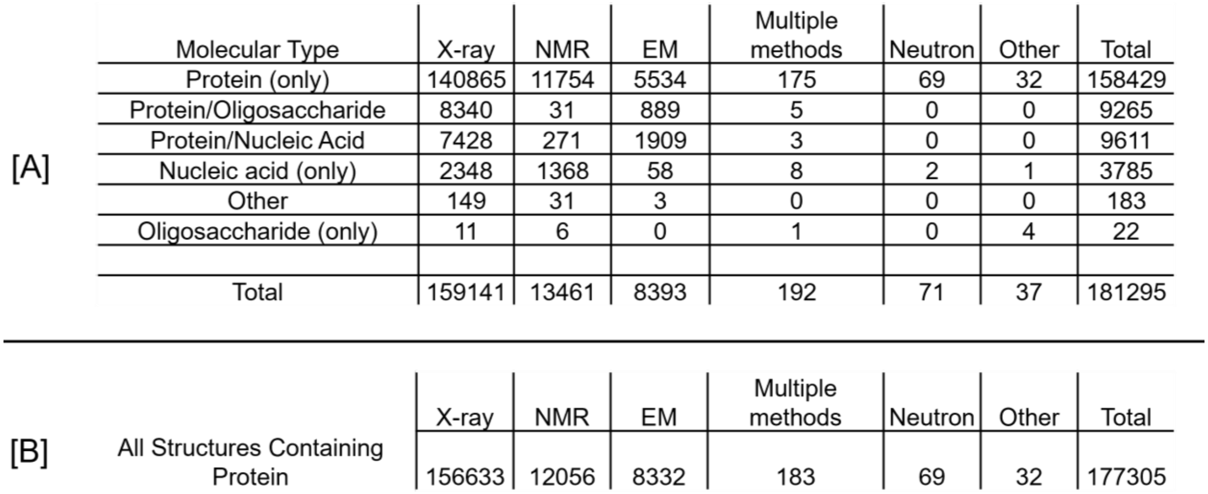


Figure 1.12 Molecular probes used in sections 2-4 by molecular weight
Molecular oxygen, MA242, and the high-affinity DNA duplex bound to PU.1 all functioned well as probes to yield structure-function insights in the 3 protein systems examined



[A]

Molecular Type	X-ray	NMR	EM	Multiple methods	Neutron	Other	Total
Protein (only)	140865	11754	5534	175	69	32	158429
Protein/Oligosaccharide	8340	31	889	5	0	0	9265
Protein/Nucleic Acid	7428	271	1909	3	0	0	9611
Nucleic acid (only)	2348	1368	58	8	2	1	3785
Other	149	31	3	0	0	0	183
Oligosaccharide (only)	11	6	0	1	0	4	22
Total	159141	13461	8393	192	71	37	181295

[B]

	X-ray	NMR	EM	Multiple methods	Neutron	Other	Total
All Structures Containing Protein	156633	12056	8332	183	69	32	177305

Table 1.1 Deposited structures in the RCSB Protein Data Bank by Structure Determination Method

[A] Number of molecular structures currently deposited in the RCSB Protein Data Bank sorted by molecular type and method utilized for structure determination **[B]** A summary of all structures containing protein chains sorted by structure determination method

1.7 References

[1] Angov E. (2011). Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology journal*, *6*(6), 650–659.

[2] Graves, P. R., & Haystead, T. A. (2002). Molecular biologist's guide to proteomics. *Microbiology and molecular biology reviews : MMBR*, *66*(1), 39–63.

[3] Kim, MS., Pinto, S., Getnet, D. *et al.* A draft map of the human proteome. *Nature* **509,** 575–581 (2014).

[4] Ponomarenko, E. A., Poverennaya, E. V., Ilgisonis, E. V., Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., Lisitsa, A. V., & Archakov, A. I. (2016). The Size of the Human Proteome: The Width and Depth. *International journal of analytical chemistry*, *2016*, 7436849.

[5] Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science. 2000 Aug 11;289(5481):920-30.

[6] Yurovsky, A., Amin, M. R., Gardin, J., Chen, Y., Skiena, S., & Futcher, B. (2018). Prokaryotic coding regions have little if any specific depletion of Shine-Dalgarno motifs. *PloS one*, *13*(8), e0202768.

[7] Ramanathan, A., Robb, G. B., & Chan, S. H. (2016). mRNA capping: biological functions and applications. *Nucleic acids research*, *44*(16), 7511–7526.

[8] Kozak M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev. 1983 Mar;47(1):1-45.

[9] Hinttala, R., Sasarman, F., Nishimura, T., Antonicka, H., Brunel-Guitton, C., Schwartzentruber, J., Fahiminiya, S., Majewski, J., Faubert, D., Ostergaard, E., Smeitink, J. A., & Shoubridge, E. A. (2015). An N-terminal formyl methionine on COX 1 is required for the assembly of cytochrome c oxidase. *Human molecular genetics*, *24*(14), 4103–4113.

[10] Varshney U, RajBhandary UL. Role of methionine and formylation of initiator tRNA in initiation of protein synthesis in Escherichia coli. J Bacteriol. 1992 Dec;174(23):7819-7826

[11] Simonović, M., & Steitz, T. A. (2009). A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochimica et biophysica acta*, *1789*(9-10), 612–623.

[12] Trobro S., Åqvist, J. Mechanism of peptide bond synthesis on the ribosome. *Proceedings of the National Academy of Sciences*. 2005, 102 (35) 12395-12400

[13] Taylor, D. J., Nilsson, J., Merrill, A. R., Andersen, G. R., Nissen, P., & Frank, J. (2007). Structures of modified eEF2 80S ribosome complexes reveal the role of GTP hydrolysis in translocation. *The EMBO journal*, *26*(9), 2421–2431.

[14] Song H, Mugnier P, Das AK, Webb HM, Evans DR, Tuite MF, Hemmings BA, Barford D. The crystal structure of human eukaryotic release factor eRF1--mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. Cell. 2000;100(3):311-21.

[15] Atkins J.F., Gesteland R. The 22nd Amino Acid. *Science.* 2002; 296(5572) 1409-1410

[16] Fu LH, Wang XF, Eyal Y, She YM, Donald LJ, Standing KG, Ben-Hayyim G. A selenoprotein in the plant kingdom. Mass spectrometry confirms that an opal codon (UGA) encodes selenocysteine in Chlamydomonas reinhardtii gluththione peroxidase. J Biol Chem. 2002 Jul 19;277(29):25983-91.

[17] Brabham R, Fascione MA. Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. Chembiochem. 2017 18(20):1973-1983.

[18] Tonino, P., Kiss, B., Strom, J. *et al.* The giant protein titin regulates the length of the striated muscle thick filament. *Nat Commun* **8,** 1041 (2017).

[19] Kemnitz C.R., Loewen M.J. “Amide Resonance” Correlates with a Breadth of C−N Rotation Barriers. *Journal of the American Chemical Society* 2007 *129* (9), 2521-2528

[20] Hollingsworth, S. A., & Karplus, P. A. (2010). A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomolecular concepts*, *1*(3-4), 271–283.

[21] Astbury, W. T., and Street, A., X-ray studies of the structure of hair, wool, and related fibres.- I. General. *Phil. Trans. Roy. Soc.*, A, 1932 **230**, 75.

[22] Pauling L., Corey R.B., Branson H.R.. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences*. 1951, 37 (4) 205-211

[23] Pauling L., Corey R.B. Two Pleated-Sheet Configurations of Polypeptide Chains Involving Both Cis and Trans Amide Groups. *Proceedings of the National Academy of Sciences*. 1953 39: 247-52

[24] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983 ;22(12):2577-637.

[25] Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical journal*, 75(1), 422–427.

[26] Zhang, B., Li, J., & Lü, Q. (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. BMC bioinformatics, 19(1), 293.

[27] Roder H. Stepwise helix formation and chain compaction during protein folding *Proceedings of the National Academy of Sciences*. 2004, 101 (7) 1793-1794

[28] Masui R, Takahata Y, Inoue M, Iio Y, Okanishi H, Kim K, Nakagawa N, Yura K, Kuramitsu S. Structural insights of post-translational modification sites in the proteome of Thermus thermophilus. J Struct Funct Genomics. 2014 15(3):137-51

[29] Standley D.M., Eyrich V.A., An Y., Pincus D.L., Gunn J.R., Friesner R.A. Protein structure prediction using a combination of sequence-based alignment, constrained energy minimization, and structural alignment. *Proteins*. 2001 5:133-9

[30] Dee, D. R., & Woodside, M. T. Comparing the energy landscapes for native folding and aggregation of PrP. *Prion*, 2016 10(3), 207–220.

[31] Basu, M. K., Poliakov, E., & Rogozin, I. B. Domain mobility in proteins: functional and evolutionary implications. *Briefings in bioinformatics*, 2009 10(3), 205–216.

[32] Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*. 1996 8;274(5289):948-53.

[33] Leslie P.L., Ke H., Zhang Y.. The MDM2 RING domain and central acidic domain play distinct roles in MDM2 protein homodimerization and MDM2-MDMX protein heterodimerization. *J Biol Chem.* 2015 15;290(20):12941-50.

[34] Yu, X., Wang, C. & Li, Y. Classification of protein quaternary structure by functional domain composition. BMC Bioinformatics. 2006 7, 187

[35] Aharoni R, Tobi D. Dynamical comparison between myoglobin and hemoglobin. *Proteins*. 2018 86(11):1176-1183

[36] Burda, P., Laslo, P. & Stopka, T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia.* 2010 24, 1249–1257

[37] Hvidsten TR, Laegreid A, Kryshtafovych A, Andersson G, Fidelis K, Komorowski J. A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity. PLoS One. 2009 4(7)

[38] Pesce, A., Couture, M., Dewilde, S., Guertin, M., Yamauchi, K., Ascenzi, P., Moens, L., & Bolognesi, M. A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family. *The EMBO journal*. 2000 19(11), 2424–2434.

[39] Barrick, D., Ho, N., Simplaceanu, V. Dahlquist, F.W., Ho, C. A test of the role of the proximal histidines in the Perutz model for cooperativity in haemoglobin. *Nat. Struct. Mol. Biol.* 1997 4, 78–83 (1997).

[40] Lüdemann, J., Verissimo, K.M., Dreger, K., Fago, A., Schneider, I., Burmester, T. Globin E is a myoglobin-related, respiratory protein highly expressed in lungfish oocytes. 2019 *Sci. Rep.* 9, 280

[41] Bucci E, Kowalczyk J. An. alpha-helical domain in hemoglobin regulated by the proximal histidine. *Biochemistry*. 1982 1;21(23):5898-901.

[42] Lennon, A. J., Scott, N. R., Chapman, B. E., & Kuchel, P. W. Hemoglobin affinity for 2,3-bisphosphoglycerate in solutions and intact erythrocytes: studies using pulsed-field gradient nuclear magnetic resonance and Monte Carlo simulations. *Biophysical journal*, 1994 67(5), 2096–2109.

[43] Dang J., Kuo M.L., Eischen C.M., Stepanova L., Sherr C.J., Roussel M.F. The RING Domain of Mdm2 Can Inhibit Cell Proliferation. Cancer Res. 2002 62(4) 1222-1230.

[44] Shloush J, Vlassov JE, Engson I, Duan S, Saridakis V, Dhe-Paganon S, Raught B, Sheng Y, Arrowsmith CH. Structural and functional comparison of the RING domains of two p53 E3 ligases, Mdm2 and Pirh2. *J. Biol. Chem*. 2011 11;286(6):4796-808.

[45] Azmi, A., Philip, P., Beck, F. et al. MI-219-zinc combination: a new paradigm in MDM2 inhibitor-based therapy. Oncogene. 2011 30, 117–126

[46] Wu W, Xu C, Ling X, et al. Targeting RING domains of Mdm2-MdmX E3 complex activates apoptotic arm of the p53 pathway in leukemia/lymphoma cells. Cell Death Dis. 2015 6(12):e2035

[47] Piette, J., Neel, H., Maréchal, V. Mdm2: keeping p53 under control. *Oncogene.* 1997 15, 1001–1010

[48] Kodandapani, R., Pio, F., Ni, CZ. et al. A new pattern for helix–turn–helix recognition revealed by the PU.l ETS–domain–DNA complex. Nature. 1996 380, 456–460

[49] Hosokawa H, Ungerbäck J, Wang X, Matsumoto M, Nakayama KI, Cohen SM, Tanaka T, Rothenberg EV. Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. Immunity. 2018 Jun 19;48(6):1119-1134.

[50] PDB Data Distribution by Experimental Method and Molecular Type. (2021, August). Retrieved from RCSB Protein Data Bank: https://www.rcsb.org/stats/summary

[51] Number of Released PDB Structures per Year. (2021, August). Retrieved from RCSB Protein Data Bank: https://www.rcsb.org/stats/all-released-structures

[52] Steno, N. (1669). De solido intra solidum naturaliter contento dissertationis prodromus. Florentiae

[53] Hauy, R. (1801). Traité de Mineralogie. Paris: French Mining Council.

[54] Nespolo, M. Direction indices for crystal lattices. J. Appl. Cryst. (2017). 50, 1541-1544

[55] Hessel JFC (1831). Kristallometrie oder Kristallonomie und Kristallographie. Leipzig.

[56] Bravais A (1850). "Mémoire sur les systèmes formés par des points distribués regulièrement sur un plan ou dans l'espace". Journal de l'École Polytechnique. 19: 1.

[57] Chuprunov, E.V. Fedorov pseudosymmetry of crystals: Review. Crystallogr. Rep. 52, 1–11 (2007).

[58] Schönflies A (1891). Kristallsysteme und Kristallstruktur. Leipzig.

[59] Barlow W. Probable nature of the internal symmetry of crystals. Nature. 1883 29 (738): 186.

[60] Rontgen W.C. Ueber eine neue Art von Strahlen. *Annalen der Physik*. 1898 300 (1) 12-17

[61] Friedrich W; Knipping P; von Laue M (1912). "Interferenz-Erscheinungen bei Röntgenstrahlen". Sitzungsberichte der Mathematisch-Physikalischen Classe der Königlich-Bayerischen Akademie der Wissenschaften zu München. 1912: 303.

[62] Bragg WL (1912). The Specular Reflexion of X-rays. Nature. 90 (2250): 410.

[63] Bragg (1913). The Structure of Some Crystals as Indicated by their Diffraction of X-rays. Proc. R. Soc. Lond. A89 (610): 248–277.

[64] Kendrew, J.C., Bodo, G., Dintzis, H. et al. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. Nature 181, 662–666 (1958)

[65] Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D. Solvent entropy contribution to the free energy of protein crystallization. Acta Crystallogr D Biol Crystallogr. 2002 58:1611-6.

[66] D. Hekmat, D. Hebel, H. Schmid, D. Weuster-Botz, Crystallization of lysozyme: From vapor diffusion experiments to batch crystallization in agitated ml-scale vessels. Process Biochemistry, 2007 42(12), 1649-1654,

[67] Shahzad Majeed, Gilad Ofek, Adam Belachew, Chih-chin Huang, Tongqing Zhou, Peter D. Kwong, Enhancing Protein Crystallization through Precipitant Synergy. Structure. 2003 11(9), 1061-1070

[68] Sang-Mok Chang, Jong-Min Kim, In-Ho Kim, Dong-Myung Shin, and Woo-Sik Kim.

Agglomeration Control of l-Ornithine Aspartate Crystals by Operating Variables in Drowning-out Crystallization. Industrial & Engineering Chemistry Research 2006 45 (5), 1631-1635

[69] Gorrec F. (2016). Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. Drug discovery today, 21(5), 819–825.

[70] Pflugrath J. W. (2015). Practical macromolecular cryocrystallography. Acta crystallographica. Section F, Structural biology communications, 71(6), 622–642.

[71] P.P. Kane, Lynn Kissel, R.H. Pratt, S.C. Roy. Elastic scattering of γ-rays and X-rays by atoms. Physics Reports. 1986 140(2), 75-159

[72] W. Kabsch, Xds, Acta Crystallographica Section D: Biological Crystallography, 66 (2010) 125-132.

[73] 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, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallographica Section D: Biological Crystallography, 66 (2010) 213-221.

[74] Wang J. (2015). Estimation of the quality of refined protein crystal structures. Protein science : a publication of the Protein Society, 24(5), 661–669.

[75] Maveyraud, L., & Mourey, L. (2020). Protein X-ray Crystallography and Drug Discovery. Molecules. 25(5), 1030.

[76] Rupp, B. (2009). Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. United States: CRC Press.

[77] Matthews B.W. X-ray Crystallographic Studies of Proteins. Annual Review of Physical Chemistry 1976 27:1, 493-493

[78] Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein science : a publication of the Protein Society, 12(9), 1865–1871.

[79] Satoh T, Samejima T, Watanabe M, Nogi S, Takahashi Y, Kaji H, Teplyakov A, Obmolova G, Kuranova I, Ishii K. Molecular cloning, expression, and site-directed mutagenesis of inorganic pyrophosphatase from Thermus thermophilus HB8. J Biochem. 1998 Jul;124(1):79-88.

[80] Slovakova M, Bilkova Z. Contemporary Enzyme-Based Methods for Recombinant Proteins In Vitro Phosphorylation. Catalysts. 2021; 11(8):1007.

[81] Berger SA, Evans PR. Site-directed mutagenesis identifies catalytic residues in the active site of Escherichia coli phosphofructokinase. Biochemistry. 1992. 31(38). 9237-42

[82] S. Aota, T. Nagai, K.M. Yamada. Characterization of regions of fibronectin besides the arginine-glycine-aspartic acid sequence required for adhesive function of the cell-binding domain using site-directed mutagenesis. Journal of Biological Chemistry. 1991 266(24), 15938-15943

[83] Roy A, Hutcheon ML, Duncan TM, Cingolani G. Improved crystallization of Escherichia coli ATP synthase catalytic complex (F1) by introducing a phosphomimetic mutation in subunit ε. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2012 68(10) 1229-33.

[84] Shi, D., & Gu, W. (2012). Dual Roles of MDM2 in the Regulation of p53: Ubiquitination Dependent and Ubiquitination Independent Mechanisms of MDM2 Repression of p53 Activity. Genes & cancer, 3(3-4), 240–248.

[85] Saville, M. K., Sparks, A., Xirodimas, D. P., Wardrop, J., Stevenson, L. F., Bourdon, J. C., Woods, Y. L., & Lane, D. P. (2004). Regulation of p53 by the ubiquitin-conjugating enzymes UbcH5B/C in vivo. The Journal of biological chemistry, 279(40), 42169–42181.

[86] Kubbutat, M. H., Jones, S. N., & Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. Nature, 387(6630), 299–303.

[87] Goldberg, Z., Vogt Sionov, R., Berger, M., Zwang, Y., Perets, R., Van Etten, R. A., Oren, M., Taya, Y., & Haupt, Y. (2002). Tyrosine phosphorylation of Mdm2 by c-Abl: implications for p53 regulation. The EMBO journal, 21(14), 3715–3727.

[88] Kastenhuber, E. R., & Lowe, S. W. (2017). Putting p53 in Context. Cell, 170(6), 1062–1078.

[89] Magnussen, H. M., Ahmed, S. F., Sibbet, G. J., Hristova, V. A., Nomura, K., Hock, A. K., Archibald, L. J., Jamieson, A. G., Fushman, D., Vousden, K. H., Weissman, A. M., & Huang, D. T. (2020). Structural basis for DNA damage-induced phosphoregulation of MDM2 RING domain. Nature communications, 11(1), 2094.

[90] Uchida, T., Gill, D. M., & Pappenheimer, A. M., Jr (1971). Mutation in the structural gene for diphtheria toxin carried by temperate phage . Nature: New biology, 233(35), 8–11.

[91] Bell, C. E., & Eisenberg, D. (1996). Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. Biochemistry, 35(4), 1137–1149.

[92] Daniels, C. C., Rogers, P. D., & Shelton, C. M. (2016). A Review of Pneumococcal Vaccines: Current Polysaccharide Vaccine Recommendations and Future Protein Antigens. The journal of pediatric pharmacology and therapeutics : JPPT : the official journal of PPAG, 21(1), 27–35.

[93] Malito, E., Bursulaya, B., Chen, C., Lo Surdo, P., Picchianti, M., Balducci, E., Biancucci, M., Brock, A., Berti, F., Bottomley, M. J., Nissum, M., Costantino, P., Rappuoli, R., & Spraggon, G. (2012). Structural basis for lack of toxicity of the diphtheria toxin mutant

CRM197. Proceedings of the National Academy of Sciences of the United States of America, 109(14), 5229–5234.

[94] Pace, C. N., Grimsley, G. R., & Scholtz, J. M. (2009). Protein ionizable groups: pK values and their contribution to protein stability and solubility. The Journal of biological chemistry, 284(20), 13285–13289.

[95] Isom D.G., Castañeda C.A., Cannon B.R., García-Moreno E B. Large shifts in pKa values of lysine residues buried inside a protein. Proceedings of the National Academy of Sciences. 2011, 108 (13) 5260-5265

[96] Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. Nature reviews. Microbiology, 9(5), 330–343

[97] Selivanov, V. A., Zeak, J. A., Roca, J., Cascante, M., Trucco, M., & Votyakova, T. V. (2008). The role of external and matrix pH in mitochondrial reactive oxygen species generation. The Journal of biological chemistry, 283(43), 29292–29300

[98] Di Russo, N. V., Estrin, D. A., Martí, M. A., & Roitberg, A. E. (2012). pH-Dependent conformational changes in proteins and their effect on experimental pK(a)s: the case of Nitrophorin 4. PLoS computational biology, 8(11), e1002761

[99] di Luccio, E., Ishida, Y., Leal, W. S., & Wilson, D. K. (2013). Crystallographic observation of pH-induced conformational changes in the Amyelois transitella pheromone-binding protein AtraPBP1. PloS one, 8(2), e53840.

[100] Thomaston, J. L., Woldeyes, R. A., Nakane, T., Yamashita, A., Tanaka, T., Koiwai, K., Brewster, A. S., Barad, B. A., Chen, Y., Lemmin, T., Uervirojnangkoorn, M., Arima, T., Kobayashi, J., Masuda, T., Suzuki, M., Sugahara, M., Sauter, N. K., Tanaka, R., Nureki, O., Tono, K., … DeGrado, W. F. (2017). XFEL structures of the influenza M2 proton channel:

Room temperature water networks and insights into proton conduction. Proceedings of the National Academy of Sciences of the United States of America, 114(51), 13357–13362

[101] Fujioka, Y., Tsuda, M., Nanbo, A., Hattori, T., Sasaki, J., Sasaki, T., Miyazaki, T., & Ohba, Y. (2013). A Ca(2+)-dependent signalling circuit regulates influenza A virus internalization and infection. Nature communications, 4, 2763

[102] Akole, A., & Warner, J. M. (2019). Model of influenza virus acidification. PloS one, 14(4), e0214448.

[103] Cady, S. D., Luo, W., Hu, F., & Hong, M. (2009). Structure and function of the influenza A M2 proton channel. Biochemistry, 48(31), 7356–7364

[104] Miao, Y., Fu, R., Zhou, H. X., & Cross, T. A. (2015). Dynamic Short Hydrogen Bonds in Histidine Tetrad of Full-Length M2 Proton Channel Reveal Tetrameric Structural Heterogeneity and Functional Mechanism. Structure, 23(12), 2300–2308.

[105] Gelenter, M. D., Mandala, V. S., Niesen, M., Sharon, D. A., Dregni, A. J., Willard, A. P., & Hong, M. (2021). Water orientation and dynamics in the closed and open influenza B virus M2 proton channels. Communications biology, 4(1), 338.

[106] Li, S., Sieben, C., Ludwig, K., Höfer, C. T., Chiantia, S., Herrmann, A., Eghiaian, F., & Schaap, I. A. (2014). pH-Controlled two-step uncoating of influenza virus. Biophysical journal, 106(7), 1447–1456.

[107] Gao, J., Gui, M., & Xiang, Y. (2020). Structural intermediates in the low pH-induced transition of influenza hemagglutinin. PLoS pathogens, 16(11), e1009062.

[108] Cady, S. D., Schmidt-Rohr, K., Wang, J., Soto, C. S., Degrado, W. F., & Hong, M. (2010). Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. Nature, 463(7281), 689–692.

[109] Liu, J., Cao, R., Xu, M., Wang, X., Zhang, H., Hu, H., Li, Y., Hu, Z., Zhong, W., & Wang, M. (2020). Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. Cell discovery, 6, 16.

[110] Rosenke, K., Jarvis, M. A., Feldmann, F., Schwarz, B., Okumura, A., Lovaglio, J., Saturday, G., Hanley, P. W., Meade-White, K., Williamson, B. N., Hansen, F., Perez-Perez, L., Leventhal, S., Tang-Huau, T. L., Nason, M., Callison, J., Haddock, E., Scott, D., Sewell, G., Bosio, C. M., … Feldmann, H. (2020). Hydroxychloroquine Proves Ineffective in Hamsters and Macaques Infected with SARS-CoV-2. bioRxiv : the preprint server for biology, 2020.06.10.145144.

[111] Hollenstein M. (2012). Nucleoside triphosphates--building blocks for the modification of nucleic acids. Molecules (Basel, Switzerland), 17(11), 13569–13591.

[112] Wang, Y., Liu, K. F., Yang, Y., Davis, I., & Liu, A. (2020). Observing 3-hydroxyanthranilate-3,4-dioxygenase in action through a crystalline lens. Proceedings of the National Academy of Sciences of the United States of America, 117(33), 19720–19730.

[113] Zhu, L., Chen, X., Abola, E. E., Jing, L., & Liu, W. (2020). Serial Crystallography for Structure-Based Drug Discovery. Trends in pharmacological sciences, 41(11), 830–839.

[114] Stefan, M. I., & Le Novère, N. (2013). Cooperative binding. PLoS computational biology, 9(6), e1003106.

[115] Perutz, M. F., Wilkinson, A. J., Paoli, M., & Dodson, G. G. (1998). The stereochemical mechanism of the cooperative effects in hemoglobin revisited. Annual review of biophysics and biomolecular structure, 27, 1–34.

[116] Balcerek, B., Steinach, M., Lichti, J., Maggioni, M. A., Becker, P. N., Labes, R., Gunga, H. C., Persson, P. B., & Fähling, M. (2020). A broad diversity in oxygen affinity to haemoglobin. Scientific reports, 10(1), 16920.

[117] Sun, K., D'Alessandro, A., Ahmed, M. H., Zhang, Y., Song, A., Ko, T. P., Nemkov, T., Reisz, J. A., Wu, H., Adebiyi, M., Peng, Z., Gong, J., Liu, H., Huang, A., Wen, Y. E., Wen, A. Q., Berka, V., Bogdanov, M. V., Abdulmalik, O., Han, L., … Xia, Y. (2017). Structural and Functional Insight of Sphingosine 1-Phosphate-Mediated Pathogenic Metabolic Reprogramming in Sickle Cell Disease. Scientific reports, 7(1), 15281.

[118] Friedmann, D., Messick, T., & Marmorstein, R. (2011). Crystallization of macromolecules. Current protocols in protein science, Chapter 17, Unit17.4.

[119] Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature, 581(7807), 215–220.

[120] Rice, L. M., & Brunger, A. T. (1999). Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly. Molecular cell, 4(1), 85–95.

[121] Shukla, A. K., Gupta, C., Srivastava, A., & Jaiman, D. (2015). Antibody fragments for stabilization and crystallization of G protein-coupled receptors and their signaling complexes. Methods in enzymology, 557, 247–258.

[122] Nomura, K., Klejnot, M., Kowalczyk, D., Hock, A. K., Sibbet, G. J., Vousden, K. H., & Huang, D. T. (2017). Structural analysis of MDM2 RING separates degradation from regulation of p53 transcription activity. Nature structural & molecular biology, 24(7), 578–587.

[123] Zaman, S., Yu, X., Bencivenga, A. F., Blanden, A. R., Liu, Y., Withers, T., Na, B., Blayney, A. J., Gilleran, J., Boothman, D. A., Loh, S. N., Kimball, S. D., & Carpizo, D. R. (2019). Combinatorial Therapy of Zinc Metallochaperones with Mutant p53 Reactivation and Diminished Copper Binding. Molecular cancer therapeutics, 18(8), 1355–1365.

[124] Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N., & Alber, T. (2011). Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proceedings of the National Academy of Sciences of the United States of America, 108(39), 16247–16252.

[125] Garman E. (2003). 'Cool' crystals: macromolecular cryocrystallography and radiation damage. Current opinion in structural biology, 13(5), 545–551.

[126] Halle B. (2004). Biomolecular cryocrystallography: structural changes during flash-cooling. Proceedings of the National Academy of Sciences of the United States of America, 101(14), 4793–4798.

[127] Frauenfelder, H., Hartmann, H., Karplus, M., Kuntz, I. D., Jr, Kuriyan, J., Parak, F., Petsko, G. A., Ringe, D., Tilton, R. F., Jr, & Connolly, M. L. (1987). Thermal expansion of a protein. Biochemistry, 26(1), 254–261.

[128] C. Savino, A.E. Miele, F. Draghi, K.A. Johnson, G. Sciara, M. Brunori, B. Vallone, Pattern of cavities in globins: the case of human hemoglobin, Biopolymers, 91 (2009) 1097-1107.

[129] R.F. Tilton, Jr., I.D. Kuntz, Jr., G.A. Petsko, Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 A, Biochemistry, 23 (1984) 2849-2857.

2 HEMOGLOBIN CRYSTALS IMMERSED IN LIQUID OXYGEN REVEAL DIFFUSION CHANNELS

Copyright © Elsevier, [Biochemical and Biophysical Research Communications, Volume 495, Issue 2, January 2018, DOI: 10.1016/j.bbrc.2017.12.038016]

2.1 Abstract

Human hemoglobin (HbA) transports molecular oxygen (O_2) from the lung to tissues where the partial pressure of O_2 is lower. O_2 binds to HbA at the heme cofactor and is stabilized by a distal histidine (HisE7). HisE7 has been observed to occupy opened and closed conformations, and is postulated to act as a gate controlling the binding/release of O_2. However, it has been suggested that HbA also contains intraprotein oxygen channels for entrances/exits far from the heme. In this study, we developed a novel method of crystal immersion in liquid oxygen prior to X-ray data collection. In the crystals immersed in liquid oxygen, the heme center was oxidized to generate aquomethemoglobin. Increases of structural flexibility were also observed in regions that are synonymous with previously postulated oxygen channels. These regions also correspond to medically relevant mutations which affect O_2 affinity. The way HbA utilizes these O_2 channels could have a profound impact on understanding the relationship of HbA O_2 transport within these disease conditions. Finally, the liquid oxygen immersion technique can be utilized as a new tool to crystallographically examine proteins and protein complexes which utilize O_2 for enzyme catalysis or transport.

2.2 Introduction

Human hemoglobin (HbA) plays a vital role in transporting oxygen (O_2) from the lungs to the tissues [1]. HbA is a tetramer consisting of two α-subunits and two β-subunits [1]. In each HbA subunit, there is a prosthetic heme where an O_2 molecule can bind at the ferrous atom [2].

The tetrameric HbA is an allosteric protein in which each subunit can switch between the tensestate (low O_2 affinity) and the relaxed-state (high O_2 affinity) [3-5]. The subunits in the quaternary assembly show a positive cooperativity for O_2 binding [5]. The O_2 pressure in the lung is relatively high so that HbA in the lungs can bind O_2 molecules (oxy-HbA) [6]. When oxy-HbA is transported to other tissues though blood circulation, O_2 molecules are released where the pressure is lower. Acidic conditions also facilitate the release of O_2 molecules from oxy-HbA [7]. Therefore, there must be pathways for O_2 migration because the heme iron center is within the HbA protein. These pathways could allow HbA to sense the environmental O_2 pressure. O_2 molecules need to enter and exit HbA through these pathways. In the O_2 binding pocket of HbA, there is a distal histidine residue (named HisE7; αH58, and βH63) that forms a hydrogen bond with the heme-bound O_2 molecule [8]. The hydrogen bond stabilizes the O_2 molecule bound in this pocket. In addition, HisE7 can be regarded as a gate to shield the heme-bound O_2 molecule from the solvent [9]. The sidechain of HisE7 can assume an open or closed conformation, which has been postulated to be a gate for O_2 to enter or exit the alleged E7 channel [10]. In a recent study by molecular dynamic simulation, however, it was shown that the mechanism of O_2 escape controlled by HisE7 may not be as what was previously postulated [11]. In fact, O_2 molecules can escape from the distal histidine pocket with the conformation of HisE7 sidechain open or closed. When the sidechain of HisE7 is protonated, it will adopt the open conformation, which allows water molecules to diffuse into the distal site, but hinders the escape of apolar gas molecules including O_2. The open conformation of HisE7 increases the volume of the distal pocket and may still increase the escape of O_2 molecules from this route, but not by a lot (around 12%). Mutations of HisE7 confirmed that the E7 channel is the main route of entering and escaping the heme-O_2 binding pocket [12]. However, it is possible that there are other pathways of O_2 migration in the HbA protein [13]. By structural determination of xenon docking sites in globins, it has been suggested that alternative pathways are possible [14, 15]. These Xe docking sites are hydrophobic cavities, confirming the availability of space within HbA for alternative pathways. To verify potential pathways that include the Xe docking sites, photolysis of oxy-HbA was monitored and rebinding of O_2 molecules could be quantitated [16]. The rebinding kinetics were affected under 25atm of xenon gas, showing a decrease of escape from the α subunit. This clearly demonstrated that there are pathways for O_2 exit at least within the α subunit that go through Xe docking sites. Consistently, molecular dynamic simulations discovered multiple O_2 entrance pathways containing the Xe docking sites, in addition to the E7 channel [13]. In another simulation study, it was shown that multiple exit pathways can be found to include Xe docking sites [17]. The path of the diffusion is dependent on the allosteric state of the HbA quaternary structure. More interestingly, it was shown that O_2 molecules may escape through internal tunnels in significant portion in all allosteric states when the distal HisE7 is in the closed conformation, further confirming the presence of alternative pathways. In order to directly visualize the possible pathways of O_2 migration, we set-up a system to determine the crystal structure of HbA in liquid O_2. The idea is to trap O_2 molecules in HbA crystals that are frozen under cryo-conditions. The crystal structures of HbA were solved for different immersion times in liquid O_2. These structures are compared with previously reported HbA structures. In addition, we also observed autoxidation of the heme-iron to aquomethemoglobin when HbA crystals were immersed in liquid O_2 for a longer period of time. Accumulation of methemoglobin will cause methemoglobinemia that can be a severe disease leading to death [18-20]. Furthermore, we observed increased flexibility in areas previously postulated to be O_2 channels, giving the first direct evidence of O_2 channel entrances/exits with the endogenous ligand of HbA.

These regions can also be correlated to medically relevant mutations which show decreased/increased O_2 affinity. Finally, the novel technique of liquid O_2 immersion can be used to crystallographically study enzymes which utilize O_2 in catalysis and examine the structure-function relationships of proteins and protein complexes which transport O_2.

2.3 Results

2.3.1 Crystal Structure of HbA Immersed in Liquid O_2

When crystals of human HbA were immersed in liquid O_2, HbA was exposed to a high concentration of O_2 while the crystal was flash-frozen at cryogenic temperature (~90 K). HbA crystals were first immersed in liquid O_2 for 20 seconds, and 1 minute, and then transferred to liquid nitrogen. X-ray diffraction data were collected under cryo temperatures and the structures were determined by molecular replacement **[see Table 2.1 for statistics]**. All refined structures in this study are in the R state. Furthermore, in order to assure the utmost of accuracy in our study, we have collected and solved a control structure [**see Table 2.2]**, taken from the same batch of crystals and frozen directly in liquid nitrogen—the standard method. However, due to higher resolution structures already deposited in the PDB, and our own control structure only being of modest resolution 2.28 Å, a conscious decision was made to utilize PDB code 2DN1 as the main structure used for comparison. For the purpose of this study, our control structure and 2DN1 are identical and the same conclusions can be drawn from both of the **structures [Figure S3 and S4]**. The rationale behind choosing 2DN1 structure lies in the accuracy of the results provided by the higher resolution of the 2DN1 structure. Under physiological conditions, the heme centers of HbA must exist in the ferrous state to bind molecular O_2 [2]. Once O_2 is bound, the HbA protein environment is highly tuned to stabilize a quasi-ferrous/ferric and O_2/superoxide resonance structure [22-24]. When O_2 is released, the active site is restored to its native ferrous state. However, in highly oxidative environments, it is possible for this process to become decoupled and HbA is left with a ferric center no longer capable of binding molecular O_2, which causes a plethora of medically relevant maladies[18-20]. Compared to a published structure cryo-cooled in liquid nitrogen (PDB CODE 2DN1, 1.25 Å resolution)[8] and our control structure (PDB CODE 6BB5), we noticed 7 significant changes in the binding mode at the iron-O_2 heme centers. Crystals of HbA immersed in liquid O_2 for 20 seconds and 1 minute showed significant oxidation at all of the heme centers **[Figure 2.1]**. In both the α/β-subunits of both crystals immersed for 20 second and 1 minute a sizeable increase in the iron-O_2 bond length (2.1-2.2 Å) and a rounding of the electron density occurred. The longer bond in the new structures are similar to aquomethemoglobin with a bond length of 2.2 Å and an identical shape in electron density[25]. Thus, we have assigned these sites with a water ligand and believe that these heme centers have been oxidized to the ferric state, unable to bind molecular O_2. This indicates that, upon immersion, liquid O_2 can penetrate the crystal and access the heme sites to carry out oxidative inactivation of the protein.

2.3.2 Presence of O_2 Channels Indicated by Increased Structural Flexibility

Compared to 2DN1 and 6BB5, both structures immersed in liquid O_2 show only slight overall structural differences **[Figure 2.7 and 2.9]**. This would imply that even under these O_2 saturated conditions, in crystallo, HbA does not undergo any large structural permutations deviating away from one of the many forms of R-state (R, R2, RR2, R3)[26]. Since HbA crystals were immersed in liquid O_2 prior to collection of X-ray diffraction data, it is expected that the channels through which the O_2 molecule is transported to and from the heme site in HbA would be occupied by O_2. Unfortunately, due to crystallographic averaging and O_2's propensity to look like a water molecule, we could not assign any new O_2 binding sites with certainty. Nonetheless, in order to investigate the possibility of intraprotein O_2 channels, the local flexibility in the structures was evaluated based on their B-factors. The rationale is that the penetration of molecular O_2 into the channels would induce flexible structural changes inside HbA. In this study, refinement of the 2DN1 structure was repeated at 1.54 Å resolution in order to compare all structures at the same resolution. B-factors and their differences were plotted by residue for structures from crystals immersed in liquid O_2 and 2DN1 **[Figure 2.2A and 2.2B]**. The major changes in flexibility are highlighted in the B-factor difference plots. Furthermore, each structure was visualized by the B-factor putty preset in PyMOL **[Figure 2.3 and 2.8]**. Based on these analyses, we found that areas with increased B-factors correspond to the locations that have been computationally validated as O_2 escape routes and molecular O_2 rebinding studies[11, 13, 16, 17]. Moreover, a recent molecular dynamics study has confirmed that the regions postulated to be O_2 channels show increased flexibility and changes in their Bfactors [17]. More specifically in the α-subunit, αAB, αCE, and αGH loops are the main regions that show increased flexibility **[Figure 2.3]**. Interestingly, the αEF loop was also found to have increased flexibility, possibly indicating the presence an additional channel **[Figure 2.3]**. The degree of structural changes in these loop regions is correlated with the time of crystals in liquid O_2. As time increases, we see increasing flexibility in these regions. This suggests that different channels in HbA have a variable capacity to transport O_2 molecules. For example, the highest Bfactor increase is in the αCE loop region on the 20 sec structure. This would allow us to propose that this channel is the most utilized channel in the α-subunit. In the 1 min structure, other channel regions start to become more flexible, most notably the αEF region. As other channels become saturated, HbA seems to adjust the molecular flexibility of the channels in order to accommodate for higher O_2 concentration. In the β-subunit, canonically, only the βEF region has been proposed as an O_2

channel [17]. Our structures show increased flexibility in this region compared to 2DN1, but the βAB and the βDE loop regions also have increased flexibility. Much like the α-subunit, as the main βEF channel becomes saturated with molecular O_2, other channels start to become more flexible in response to the higher O_2 concentration. These results could indicate that the channels are modulated in order for optimal O_2 flow in and out of HbA. The increased flexibility of these channels, which could be regarded as opening the channels, is, in this case, specific for HbA's native ligand O_2 at high O_2 concentrations. We believe that the slight difference in the freezing cryo-temperature between O_2 and N2 (only around 13 K) does not play a major role in the observed increased flexibility. Furthermore, the cryo-stream at the beamline is kept at 100 K, which is 10 K higher than the temperature of liquid O_2. If temperature played a significant role, similar features should also be noted in both of the control structures. While the specific mechanism of this channel opening is outside the scope of this study, it is most likely that the opening is finely tuned to the size and electronic properties of O_2, which is why this is not seen in the structures cryo-protected in liquid nitrogen. Furthermore, upon immersion in liquid O_2 we do not see any secondary oxidation of the protein indicating that the specificity of the channels for the native ligand is highly controlled. Upon transfer back into liquid nitrogen for further cryo-storage and data collection, it is likely that O_2 may diffuse out of the channels. During this diffusion process, liquid nitrogen might be able to fill these gaps, trapping the rearranged portions of the structure revealing the footprint of the channels that were opened in liquid O_2. Furthermore, it is important to note that this increased flexibility is not noticed in the comparative structures of ferric methemoglobin, but increased flexibility was found in similar regions of the Xe incorporated hemoglobin structures (data not shown). In addition, the overall B-factor trends between 2DN1 (crystallographically a dimer), our structures (crystallographically a tetramer), and other published structures that were crystallographically tetramers (1BBB and 2W6V) were analyzed to estimate the effects of crystal packing on structural flexibility. Based on the correlation coefficient between the datasets, the alpha subunit in our liquid O_2 immersed structures is most closely correlated to that in the 2DN1 control structure, and the beta subunit showed similar correlations throughout all of the structures **[Figure 2.10]**. Therefore, the observed increased flexibility of the protein at proposed O_2 channel entrances/exits points could not be related to crystal packing.

2.3.3 Postulated O_2 Channels and Cavities

Using the program Mole 2.0 the tunnels (or channels), pores and cavities were simulated and only channels that were connected, or in close proximity, to the heme were kept and visualized **[Figure 2.4]** [27]. It is important to note that the co-crystallization ligand of toluene in 2DN1 was removed before this simulation. A probe radius of 2.5 Å, an Interior Threshold of 1.00, a Bottleneck Radius of 1.02 Å, and a Bottleneck Length of 2.52 Å were used in the simulation to locate channels. Compared to the control structure we see an increase in the channel branching connected to the heme and an increase in the number of channel entrances/exits. For our purposes of analysis, an entrance/exit was defined as a channel that was leading out of the protein and into the solvent. In the control structure, the number of entrances/exits in the α/β-subunit was 4 and 7, respectively. The 20 sec and 1 min structures contained 9 and 10 entrance/exits in the α-subunit, respectively, whereas both the β-subunit structures contained 8 entrance/exits. Besides an increase in entrances/exits, it was also noted that the diameter of the tunnels in the O_2 immersed structures appears to have increased. Finally, all the loop regions that showed increased flexibility contain a channel connected to the heme, indicated by the colored boxes in Figure 2.4. In addition, a larger cavity volume of 17,363.32 Å3 and 16,711.02 Å3 are found, respectively, in the 20 sec and 1 min structures, whereas a cavity volume of 16,096.22 Å3 is found in 2DN1 (as calculated by Mole2.0 using the abovementioned criteria). This would be indicative that the channels in HbA were opened by the liquid O_2 in crystallo. Furthermore, calculations of Solvent Accessible Surface Area (SASA) (performed with PDBePISA http://www.ebi.ac.uk/pdbe/pisa/) were carried out. There is a SASA of 23,950 Å2 , and 23,670 Å 2 in the 20 sec and 1 min structures, respectively, and a SASA of 24,710 Å2 in the 2DN1 control. This decrease in the SASA is consistent with the previous data that the hydrophobic channels present in HbA were opened to accommodate the saturating O_2 concentration.

2.4 Discussion

Immersing HbA crystals in liquid O_2 offers a unique opportunity to study HbA in a saturated O_2 environment at a cryo temperature—thus keeping the protein from disruptive oxidative damage beyond the oxidative conversion of the heme center from ferrous to the ferric form. Our experimental results clearly show that liquid O_2 can enter the crystal and oxidize the heme centers producing aquomethemoglobin. Furthermore, this confirms that the liquid O_2 can penetrate the protein in a biologically meaningful way. The B-factor analysis of the refined structures of crystals immersed in liquid O_2 for different lengths of time clearly correlates the increased flexible regions to computationally identified O_2 entrance/exit channels. Our results provided experimental evidence for the presence of these channels. Our data could also indicate that these channels are coordinated to adjust the capacity of transporting O_2 in and out of HbA protein. Finally, using Mole 2.0 we have simulated the possible channels which could be used by O_2, and have assessed the cavity size and SASA. The analysis confirmed that HbA uses its O_2 channels to modulate the accessibility of O_2 into and out of the heme site and that the channels are throughout the entire tertiary structure of the α/β-subunits. If O_2 channels regulate O_2 binding and release, this mechanism may have effects on certain disease states. Using the HbVar database, **[Table 2.4]** we searched for mutations that resulted in a change of O_2 affinity—either increased or decreased—in the O_2 entrance/exit channels. There are 12 and 32 such mutations in the α-subunit and the β-subunit, respectively. Furthermore, there are numerous mutations which resulted in a medically relevant phenotype, mostly resulting in certain types of thalassemia. These findings emphasize the functional importance of the O_2 channels. In this report we have experimentally validated the presence of O_2 channels and their structural relationship to O_2 transport in HbA, which allows further functional studies to be explored for medical implications. Also, the novel O_2 immersion technique developed in this study can be utilized by the greater scientific community as an important tool to examine proteins and enzymes that utilize O_2 as their native substrate. Most notably as a method for in crystallo rapid freeze quench, to potentially trap enzymatic intermediates that use O_2, and in other proteins which transport O_2.

2.5 Methods

2.5.1 Protein Purification and Crystallization

Human hemoglobin was isolated from whole blood and crystallized following the method described by Perutz *et. al* [21]. Whole blood was extracted using 6mL EDTA/k2 vacuette tubes. The extracted blood was centrifuged at 5,000 rpm and the supernatant was removed and discarded. The pelleted red blood cells were washed via resuspension and pelleting in a 0.9% NaCl solution (w/w) four times followed by a single wash with a 1.0% NaCl solution. The cells were lysed osmotically by doubling the volume of the sample with double-deionized water followed by a 15-minute rest period. The cell lysate was spun at 17,000 rpm in a Beckman JA-

25.50 rotor for one hour and the supernatant removed and spun for an additional 30 minutes at 17,000 rpm. The supernatant was then collected and purified using a GE Life Sciences HiPrep 26/10 desalting column that had been pre-equilibrated with a 10mM ammonium phosphate buffer at a pH of 7.0. The eluate containing hemoglobin was dialyzed against a 1.6M phosphate buffer that was prepared using a stock comprised of a 5:7 ratio of 4M monobasic sodium phosphate to 4M dibasic potassium phosphate with a final pH of 6.7. The dialyzed protein was concentrated to 40mg/mL using an Amicon Ultra-15 centrifugal concentrator. The final concentration was determined using the known extinction coefficient of 524280cm^{-1}/M [2] at 414nm in the Soret region of oxyhemoglobin. 100 µL of toluene was added to each of 4 samples that contain the components described in [**Table 2.3]**. Crystals grown at 4 °C appeared after five to seven days and were allowed to grow for three weeks prior to experiments of immersing crystals in liquid oxygen **[Figure 2.5]**. Hemoglobin crystals of approximately the same size were selected and transferred to a cryoprotectant solution comprised of 20% glycerol in crystallization mother liquor. Looped crystals were then immersed in the liquid oxygen for 20 seconds, or 1 minute and rapidly transferred to a uni-puck submerged in liquid nitrogen.

2.5.2 Generation of liquid oxygen

Liquid oxygen was generated using a cold finger apparatus submerged in liquid nitrogen [**Figure 2.6]** that was constructed following a gas condensation protocol described in Vogel's Practical Organic Chemistry [29]. The entire apparatus was purged with oxygen gas to remove contaminant gases and moisture prior to oxygen condensation. Oxygen gas was channeled through the cold finger, into a one-way gas valve to prevent backflow due to pressure drop during liquid nitrogen cooling, and out of a gas bubbler containing water to observe gas flow rates. The condensed oxygen was then transferred into a glass dewar that had been pre-chilled using liquid nitrogen. **It should be noted that liquid oxygen is extremely dangerous, should be handled with great care, and used only with clean glass containers. Contact with organic compounds or metals can result in large fires or explosion.**

2.5.3 Crystal Structure

X-ray diffraction data were collected at SER-CAT at the Advanced Photon Source, USA on the ID-22 beamline. Data was processed using the XDS package[30]. Molecular replacement was performed with the PHENIX suite [31] using maximum-likelihood procedures in PHASER and the coordinates from a HbA structure (PDB Code 2DN1, downloaded from the RCSB protein data bank) [8] as the search model. Structural refinement was carried out using phenix.refine. Modeling coordinates in electron density was performed with Coot [33]. A summary of crystallographic data is shown in [**Table 2.1 and 2.2]**. Figure preparation was carried out using PyMOL (v 1.8, Schrodinger) [34]. Theoretical channels were calculated using the Mole2.0 GUI [27] and visualized in PyMOL. All of the parameters were left at the default except for the Probe Radius (2.5), the Interior Threshold (1.0), the Bottleneck Radius (1.02), and the Bottleneck Length (2.52). All cavities, tunnels, and pores were automatically calculated using the abovementioned settings.

2.6 Figures

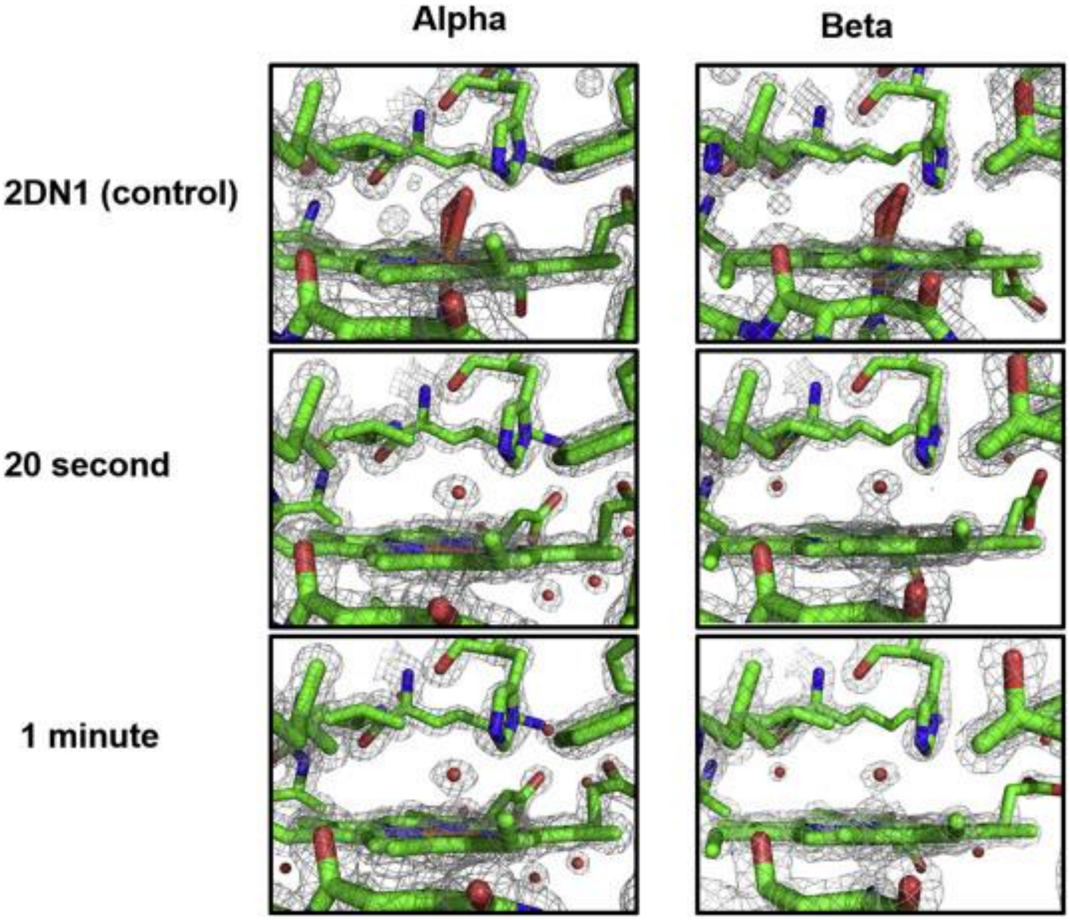


Figure 2.1 2mF0-DFc map at 2σn
2mFo-DFc map reconstruction modeled at 2σ of the alpha and beta subunit active sites, in increasing liquid O_2 immersion times, with heme and O_2/water bound to the heme iron.

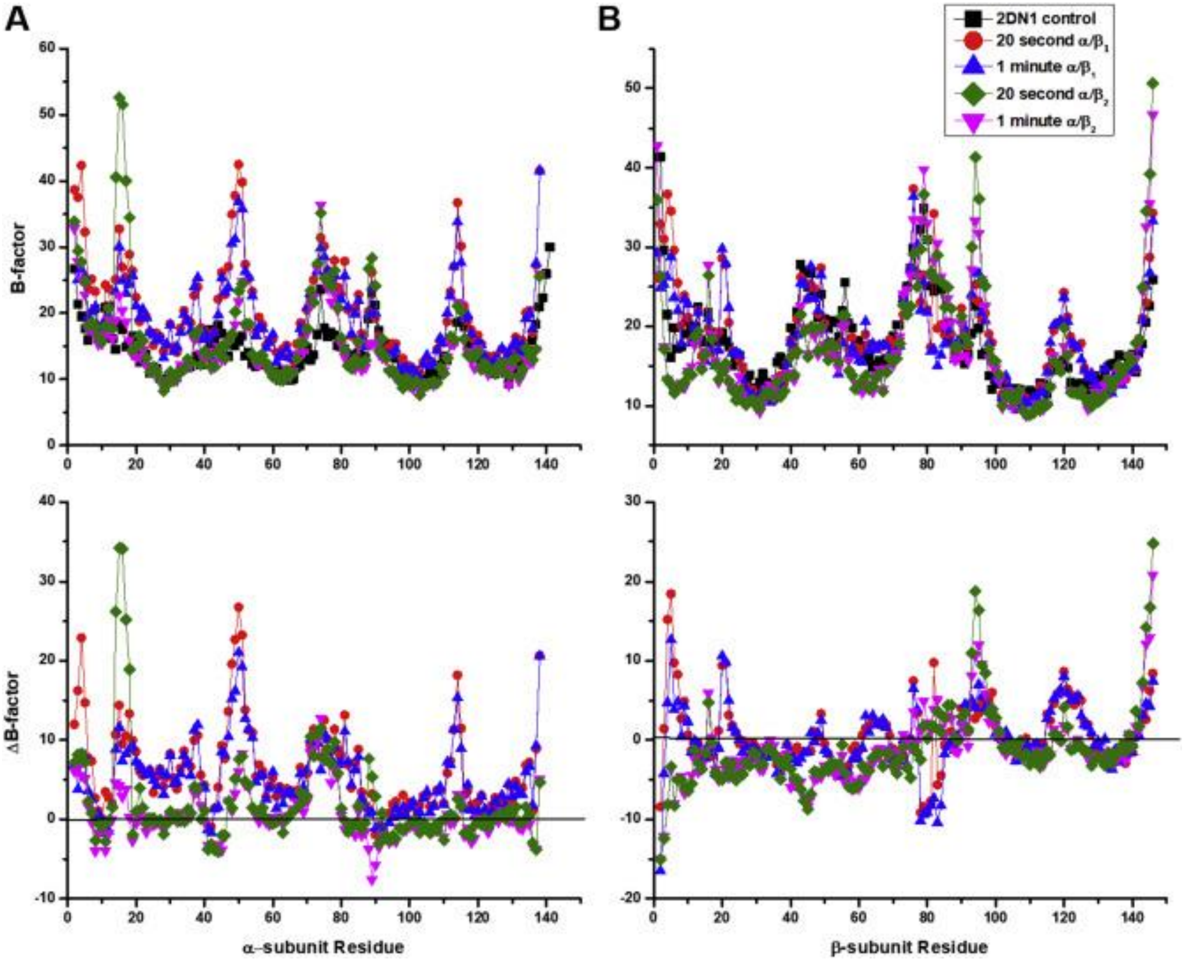


Figure 2.2 B-factor vs. residue plots for 2DN1 and liquid O_2 soaked structures
The top panel are plots of raw B-factors, while the bottom panel is a difference plot of the B-factor at each residue subtracted by the respective residue for 2DN1. The figure legend represents the structure that was plotted for either the α or β subunits depending on its respective panel (either A or B). (**A**) B-factor/ΔB-factor plots for the α-subunits. (**B**) B-factor/ΔB-factor plots for the β-subunits.

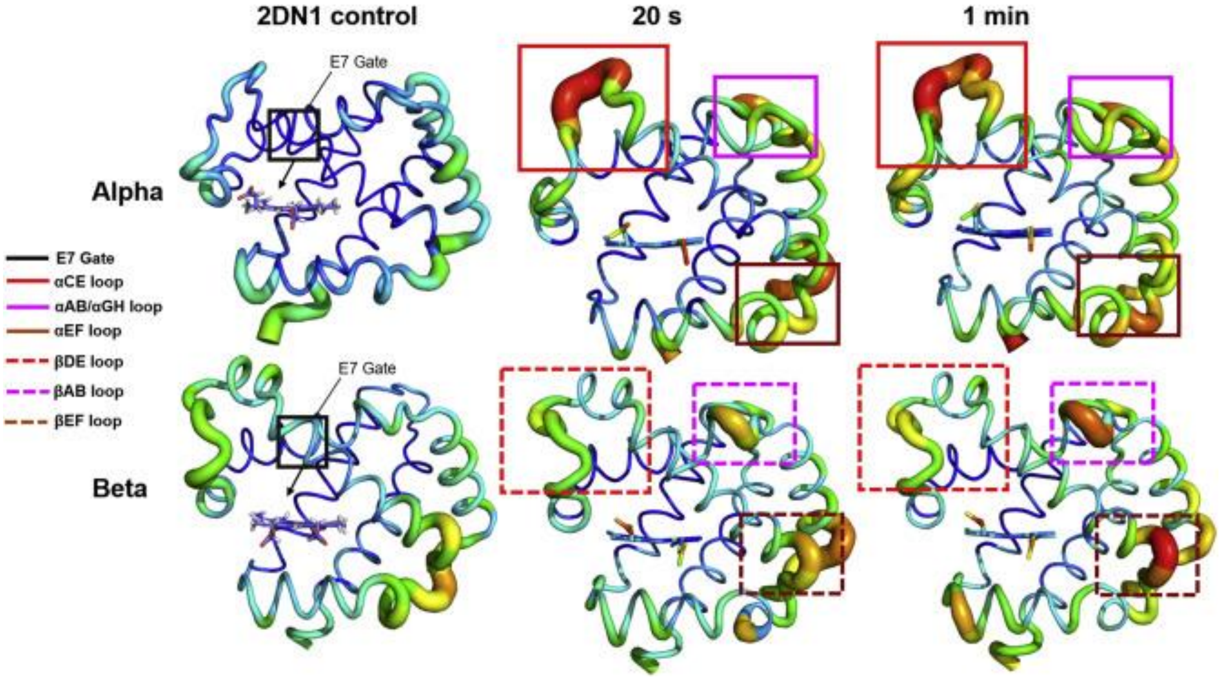


Figure 2.3 B-factor putty reconstruction (PyMOL) of the alpha and beta strands, respectively, portraying different times of HbA crystals immersed in liquid O_2
The color gradient with blue representing low B-factors and red representing higher B-factors. When crystals were immersed in liquid O_2, increased B-factors are noted in the regions of the proposed O_2 escape tunnels, with very little change in the E7(α,β) gate. Each box corresponds to the areas of higher B-factors that have been implicated in computational studies as O_2 channels. The black box signifies the E7-gate, the red box signifies the αCE/βDE loop, the magenta box signifies the αAB/αGH/βAB loop, and the brown box signifies the αEF/βEF loop. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

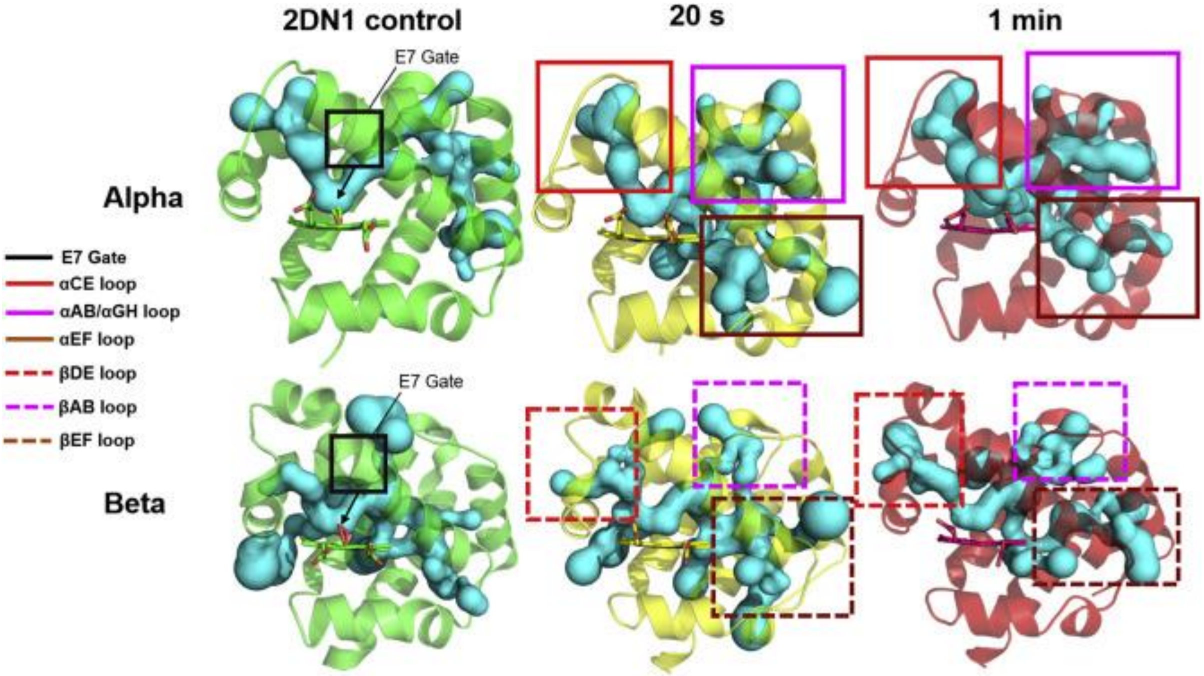


Figure 2.4 Calculated channels constructed with Mole 2.0 GUI and visualized in PyMOL Each panel represents a different time point in liquid O_2 immersion. All channels are visualized in cyan and were calculated using the same parameters in Mole 2.0. Furthermore, channels were removed from viewing if they did not contact the heme center, or lead to a channel that came into contact with the heme center, as these were deemed biologically insignificant. Each box corresponds to the areas of higher B-factors that have been implicated in computational studies as O_2 channels and also the possible entrance/exit channels in our liquid O_2 immersed structures. The black box signifies the E7-gate, the red box signifies the αCE/βDE loop, the magenta box signifies the αAB/αGH/βAB loop, and the brown box signifies the αEF/βEF loop. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

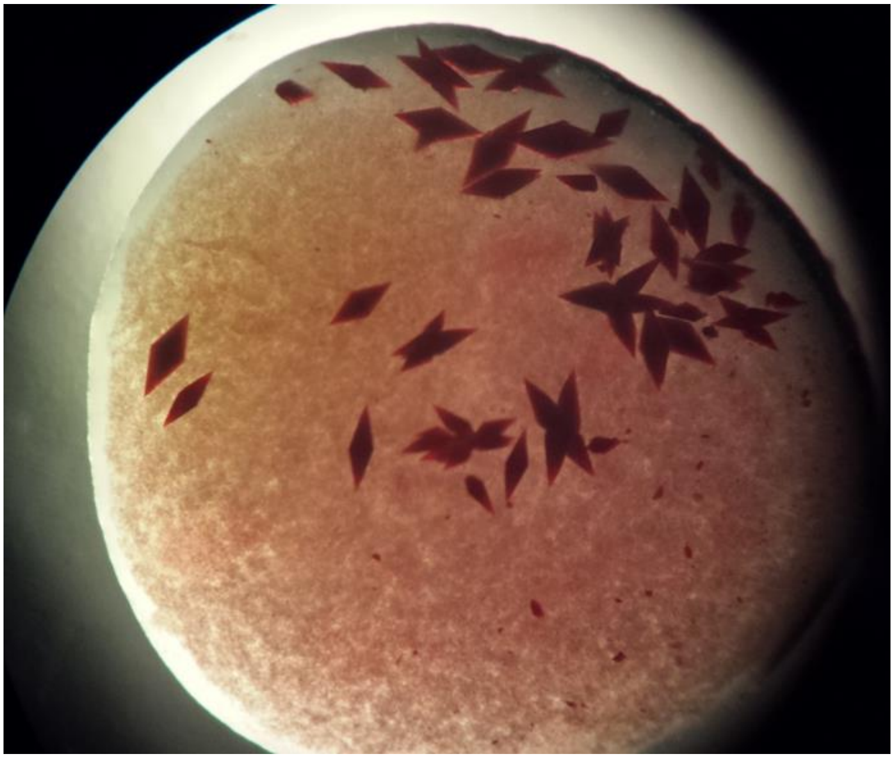


Figure 2.5 Oxyhemoglobin crystals following a 21 day growth period

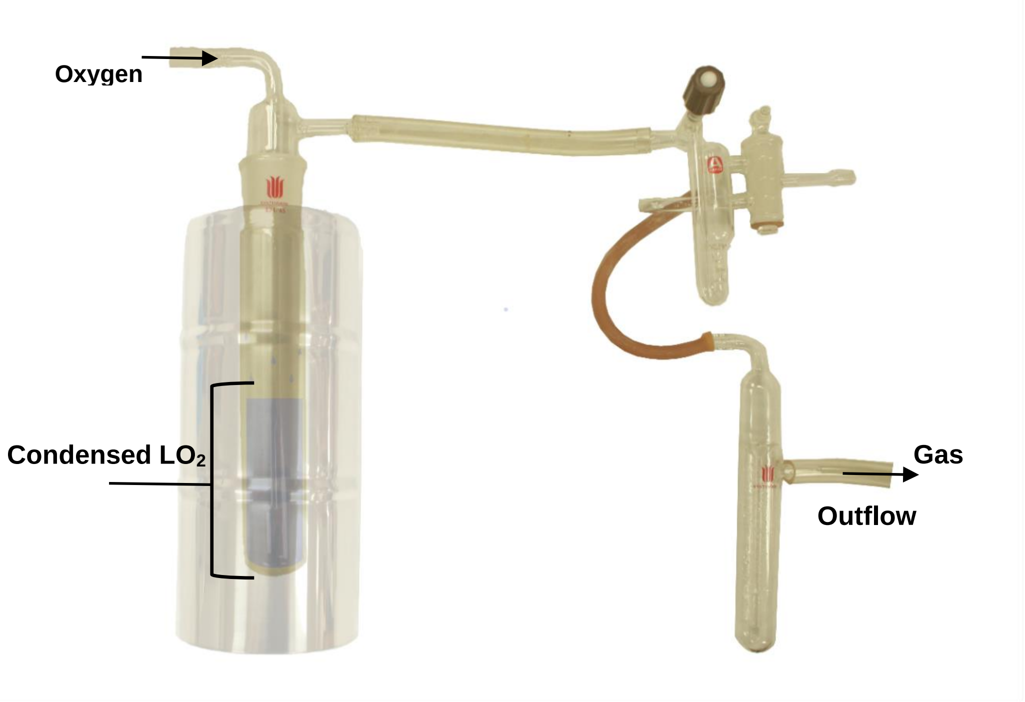


Figure 2.6 The apparatus used to generate liquid oxygen
High purity oxygen gas was slowly fed into a cold-finger trap submerged in a dewar containing liquid nitrogen. Uncondensed oxygen was channeled through a 1-way ball valve. The flow then continued into a gas bubbler filled with H_2O to monitor and optimize gas flow rates for O_2 condensation. Gas outflow was then vented into a fume hood.

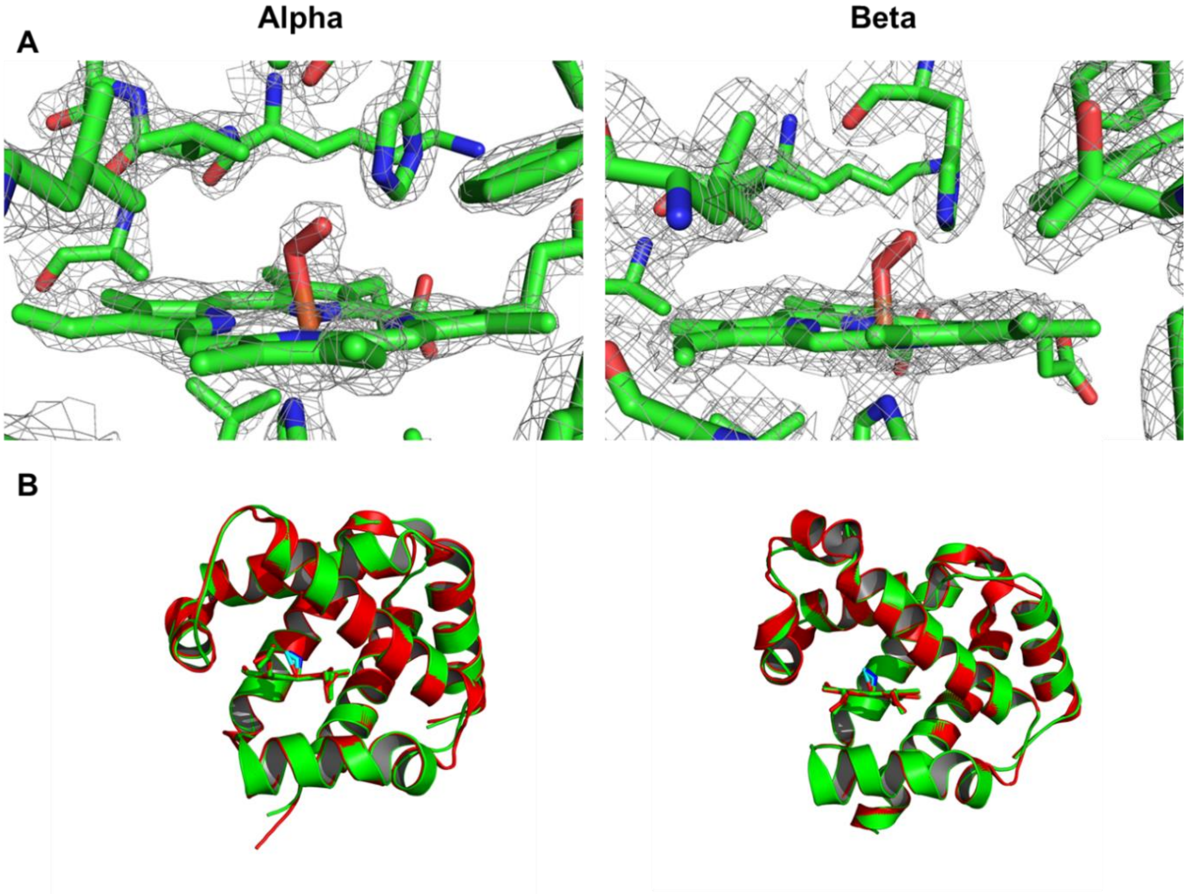


Figure 2.7 (A) 2mFo-DFc map modeled at 2σ showing the control structure heme in the R-state with oxygen bound to a ferrous center in both the α/β-subunits respectively. (B) A structural alignment of our control structure and 2DN1, showing that both structures are in the R-state with little to no change in the global structures

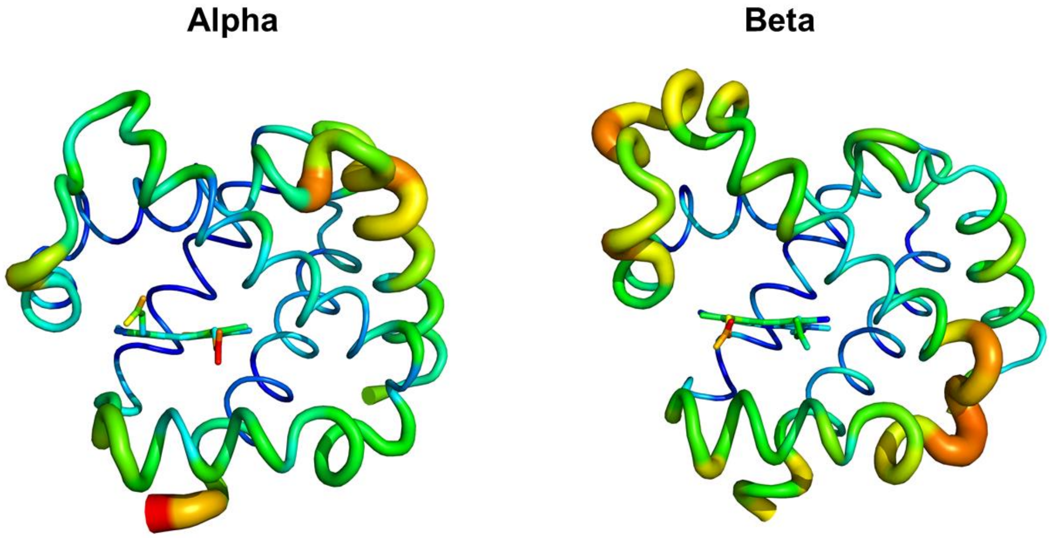


Figure 2.8 B-factor putty reconstruction of our control structure, with lower B-factors presented in blue and higher B-factors gradually going to red
The control structure, due to the lower resolution, show overall higher b-factors and similar trends as 2DN1. However, it is to be expected that there may be some minor difference due to the larger inaccuracies in our control structure rather than the more accurate 2DN1 structure.

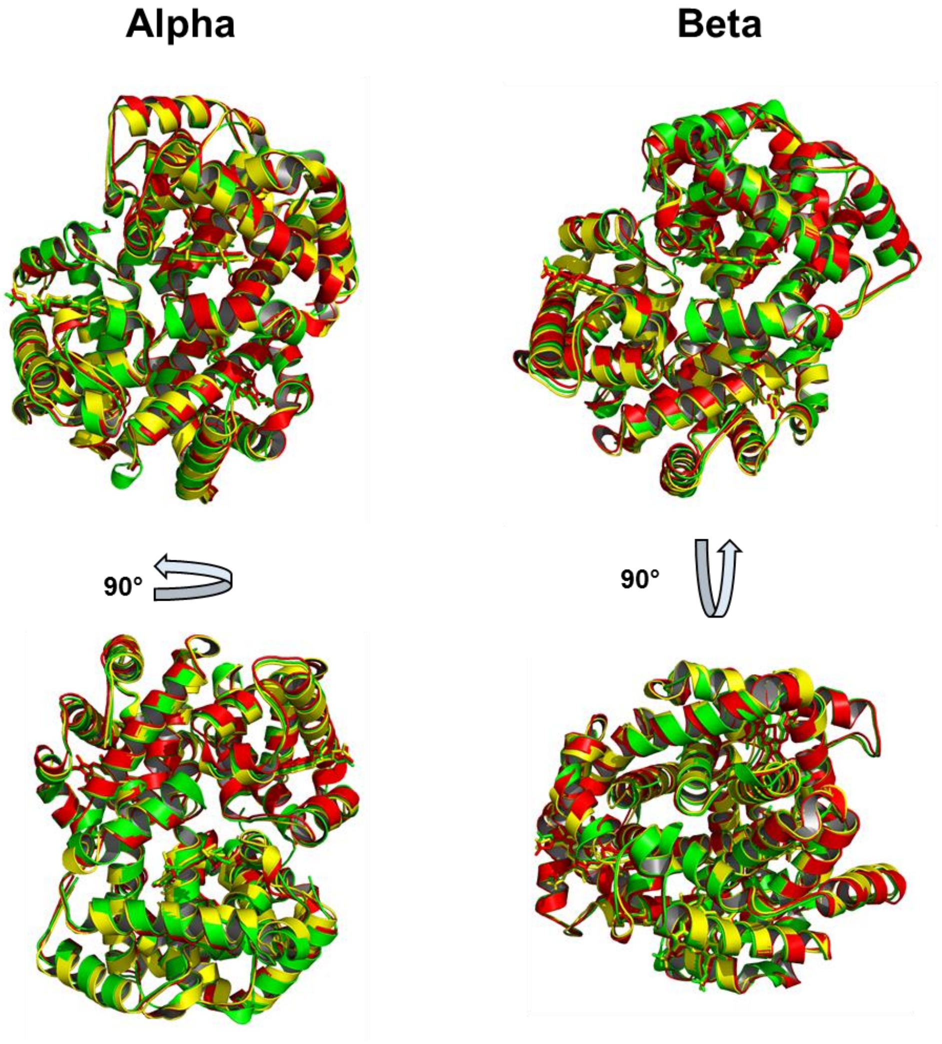


Figure 2.9 Cartoon representation and alignment of the α/β-subunits of all the HbA structures solved in this paper
(PDB: 2DN1 (green), 20 seconds (yellow), and 1 minute (red)). Only slight structural fluctuations are noticed throughout the structures, which do not lead to any noticeable deviations from the normal Oxy-R state.

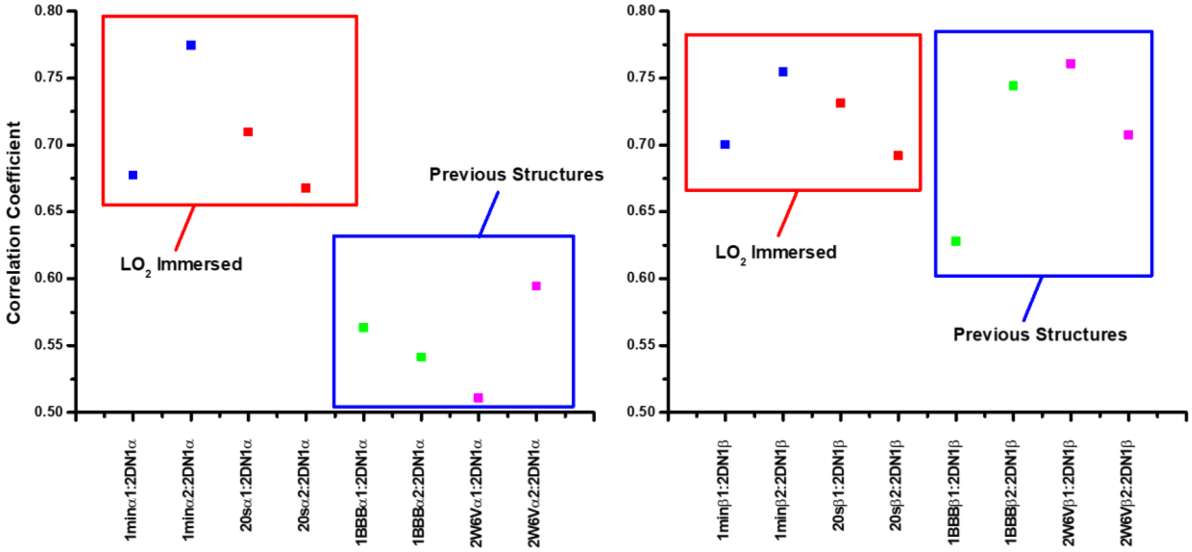


Figure 2.10 Correlation coefficients comparing the B-factor between various structures—LO_2 immersed both 1 minute and 20 seconds, 2DN1, 1BBB (R2-state and tetramer), and 2W6V (deoxyHbA with Xe)

The left panel is the alpha-subunit of the 1 minute, 20 seconds, 1BBB, and 2W6V compared to the 2DN1 control structure alpha-subunit. The right panel is beta-subunit with the same comparisons. Higher correlation coefficients indicate a positive correlation between the B-factor trends, while lower correlation coefficients indicate less correlation between the two structures, with 0 being no correlation. The left panel shows that the alpha-subunit of LO_2 are more closely correlated with 2DN1 and right panel shows that the beta-subunit is uniformly correlated throughout all of the structures, indicated that the B-factor differences are not due to crystal packing.

Table 2.1 Crystallographic Data and Refinement Statistics

	20 seconds*	**1 minute***
Wavelength	1.00 Å	1.00 Å
Resolution range	36.59 - 1.58 (1.637 - 1.58)	41.44 - 1.54 (1.595 - 1.54)
Space group	P 21 21 21	P 21 21 21
Unit cell	75.43 83.12 83.68 90 90 90	75.36 82.89 83.6 90 90 90
Total reflections	1034033 (103394)	1094635 (108135)
Unique reflections	72637 (6139)	78032 (6161)
Multiplicity	14.2 (14.4)	14.0 (14.0)
Completeness (%)	94.98 (85.41)	91.59 (79.97)
Mean I/sigma(I)	23.35 (6.66)	21.03 (5.69)
Wilson B-factor	14.98	14.33
R-merge	0.08153 (0.5652)	0.08686 (0.6431)
R-meas	0.08462 (0.5857)	0.09019 (0.6675)
R-pim	0.02244 (0.1526)	0.02405 (0.1775)
CC1/2	0.999 (0.967)	0.999 (0.959)
CC*	1 (0.992)	1 (0.989)
Reflections used in refinement	69023 (6139)	71475 (6161)
Reflections used for R-free	1996 (179)	1987 (170)
R-work	0.1824 (0.1984)	0.1843 (0.2127)
R-free	0.2071 (0.2183)	0.2066 (0.2374)
CC(work)	0.953 (0.886)	0.952 (0.868)
CC(free)	0.933 (0.830)	0.952 (0.818)
Number of non-hydrogen atom	5184	5246
macromolecules	4310	4297
ligands	172	172
solvent	702	777
Protein residues	570	570
RMS(bonds)	0.007	0.007
RMS(angles)	0.9	0.88
Ramachandran favored (%)	98.75	98.75
Ramachandran allowed (%)	1.25	1.25
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	1.32	1.33
Clashscore	4.95	6.42
Average B-factor	21.86	21.3
macromolecules	20.76	19.78
ligands	17.76	18.18
solvent	29.62	30.37

Table 2.2 Crystallographic Data and Refinement Statistics of Our Control

	6BB5
Wavelength (Å)	1
Resolution range (Å)	51.75 - 2.28 (2.362 - 2.28)
Space group	P 41 21 2
Unit cell	53.73 53.73 192.29 90 90 90
Total reflections	366351 (34841)
Unique reflections	13363 (1200)
Multiplicity	27.4 (27.3)
Completeness (%)	95.88 (89.75)
Mean I/sigma(I)	38.60 (7.04)
Wilson B-factor	27.27
R-merge	0.09305 (0.5814)
R-meas	0.09479 (0.5924)
R-pim	0.01785 (0.1112)
CC1/2	0.999 (0.955)
CC*	1 (0.988)
Reflections used in refinement	13116 (1200)
Reflections used for R-free	1312 (120)
R-work	0.2050 (0.2236)
R-free	0.2423 (0.2690)
CC(work)	0.939 (0.890)
CC(free)	0.909 (0.828)
Number of non-hydrogen atoms	2359
macromolecules	2160
ligands	90
solvent	109
Protein residues	286
RMS(bonds)	0.006
RMS(angles)	0.66
Ramachandran favored (%)	97.14
Ramachandran allowed (%)	2.14
Ramachandran outliers (%)	0.71
Rotamer outliers (%)	1.76
Clashscore	4.49
Average B-factor	30
macromolecules	29.75
ligands	32.09
solvent	33.24

Table 2.3 Crystallization conditions recreated from the paper of Perutz et. al.

Tube Number	Final molarity of mother liquor	mL 4 M Phosphate buffer stock	mL DDI H_2O	mL of 40 mg/mL Hb in 1.6M comb. phosphate buffer
1	2.65 M	2.25 mL	0.75 mL	1.0 mL
2	2.55 M	2.15 mL	0.85 mL	1.0 mL
3	2.45 M	2.05 mL	0.95 mL	1.0 mL
4	2.35 M	1.95 mL	1.05 mL	1.0 mL

*Crystals appeared after 5-7 days and were harvested after 21 days for soaking experiments.

Table 2.4 HbA Variants with Changed O_2 Affinity in Entrance/Exit Regions

Alpha-subunit			
Hb type	**Mutation**	**O_2 affinity**	**Clinical**
Hb Evanston	Trp14-R	increased	thalessemia
Hb Harbin	Lys16-Met	slightly increased	thalessemia
Hb Milledgeville	44Pro-Leu	Increased	mild Erythrocytosis
Hb Kawachi	44Pro-Arg	Increased	normal
Hb Fort de France	45His-Arg	Increased	Normal
Hb Poitiers	45His-Asp	Increased	Normal
Hb Beilinson	47Asp-Gly	Increased	Normal
Hb Chapel-Hill	74Asp-Gly	Increased	mild anemia
Hb Conakry	80Leu-Val	Decreased	marked anemia
Hb Toulon	77Pro-His	Increased	Normal
Hb Tonosho	110Ala-Thr	slightly increased	Normal
Hb Hopkins-II	112His-Asp	increased (in vitro)	Normal
Beta-subunit			
Hb type	**Mutation**	**O_2 affinity**	**Clinical**
Hb Olympia	20Val-Met	greatly increased	marked Erythrocytosis
Hb Trollhättan	20Val-Glu	increased	mild Erythrocytosis
Hb Uxbridge	20Val-Gly	slightly increased	normal
Hb Connecticut	21Asp-Gly	decreased	normal
Hb Strasbourg	23Val-Asp	increased	moderate Erythrocytosis
Hb Miyashiro	23Val-Gly	increased	normal
Hb Palmerston North	23Val-Phe	increased	Erythrocytosis, mild anemia
Hb Riverdale-Bronx	24Gly-Arg	increased	hemolytic mild anemia
Hb Moscva	24Gly-Asp	slightly decreased	hemolytic anemia,Reticulocytosis
Hb J-Auckland	25Gly-Asp	decreased	normal
Hb S-Antilles	23Val-Ile, 6Glu-val	decrased	mild anemia, Splenomegaly, Reticulocytosis
Hb Zoeterwoude	23Val-Ala	increased	marked Erythrocytosis
Hb Howden	20Val-Ala	increased	familial polycythemia
Hb Headington	72Ser-Arg	increased	mild Erythrocytosis
Hb G-Accra	73Asp-Asn	Decreased	normal
Hb Shepherds Bush	74Gly-Asp	increased	hemolytic Anemia
Hb Pasadena	75Leu-Arg	decreased	hemolytic compensated anemia
Hb Calais	76Ala-Pro	decreased	chornic anemia, severe Fe deficiencies
Hb Tigraye	79Asp-His	increased	normal
Hb Baylor	81Leu-Arg	increased	hemolytic anemia, Heinz bodies, high Hb and Red blood cell count
Hb La Roche-sur-Yon	81Leu-His	slightly increased	normal
Hb Tsurumai	82Lys-Glu	increased	erythrocytosis
Hb Hammersmith	42Phe-Ser	decreased	hemolytic severe Anemia, Heinz bodies
Hb Louisville	42Phe-Leu	decreased	hemolytic mild anemia, Heinz bodies, Reticulocytosis
Hb Sendagi	42Phe-Val	decreased	hemolytic moderate anemia, Heinz bodies, Reticulocytosis
Hb Cheverly	45Phe-Ser	decreased	hemolytic mild chronic anemia, Heinz bodies, Reticulocytosis
Hb Arta	45Phe-Cys	decreased	normal
Hb Okaloosa	48Leu-Arg	decreased	normal
Hb Willamette	51Pro-Arg	increased	normal
Hb Jacksonville	54Val-Asp	increased	hemolytic Anemia, Heinz bodies, Reticulocytosis;mainly during infections
Hb Poissy	56Gly-Arg,86Ala-Pro	increased	Erythrocytosis
Hb Muravera	47Asp-Val	increased	normal

2.7 Referrences

[1] H. Mairbaurl, R.E. Weber, Oxygen transport by hemoglobin, Comprehensive Physiology, 2 (2012) 1463-1489.

[2] M.C. Marden, N. Griffon, C. Poyart, Oxygen delivery and autoxidation of hemoglobin, Transfus Clin Biol, 2 (1995) 473-480.

[3] M.F. Perutz, Structure and mechanism of haemoglobin, British medical bulletin, 32 (1976) 195-208.

[4] J.M. Baldwin, A model of co-operative oxygen binding to haemoglobin, British medical bulletin, 32 (1976) 213-218.

[5] A. Bellelli, Hemoglobin and cooperativity: Experiments and theories, Current protein & peptide science, 11 (2010) 2-36.

[6] C.A. Finch, C. Lenfant, Oxygen transport in man, The New England journal of medicine, 286 (1972) 407-415.

[7] J.M. Salhany, D.H. Mathers, R.S. Eliot, Molecular basis for oxygen transport. Hemoglobin function and controlling factors, Advances in cardiology, 9 (1973) 53-67.

[8] S.Y. Park, T. Yokoyama, N. Shibayama, Y. Shiro, J.R. Tame, 1.25 A resolution crystal structures of human haemoglobin in the oxy, deoxy and carbonmonoxy forms, J Mol Biol, 360 (2006) 690-701.

[9] M.D. Salter, G.C. Blouin, J. Soman, E.W. Singleton, S. Dewilde, L. Moens, A. Pesce, M. Nardini, M. Bolognesi, J.S. Olson, Determination of ligand pathways in globins: apolar tunnels versus polar gates, The Journal of biological chemistry, 287 (2012) 33163-33178.

[10] I. Birukou, R.L. Schweers, J.S. Olson, Distal histidine stabilizes bound O2 and acts as a gate for ligand entry in both subunits of adult human hemoglobin, The Journal of biological chemistry, 285 (2010) 8840-8854.

[11] M.S. Shadrina, G.H. Peslherbe, A.M. English, O2 and Water Migration Pathways between the Solvent and Heme Pockets of Hemoglobin with Open and Closed Conformations of the Distal HisE7, Biochemistry, 54 (2015) 5279-5289.

[12] I. Birukou, J. Soman, J.S. Olson, Blocking the gate to ligand entry in human hemoglobin, The Journal of biological chemistry, 286 (2011) 10515-10529.

[13] M.F. Lucas, V. Guallar, Single vs. multiple ligand pathways in globins: a computational view, Biochimica et biophysica acta, 1834 (2013) 1739-1743.

[14] C. Savino, A.E. Miele, F. Draghi, K.A. Johnson, G. Sciara, M. Brunori, B. Vallone, Pattern of cavities in globins: the case of human hemoglobin, Biopolymers, 91 (2009) 1097-1107.

[15] R.F. Tilton, Jr., I.D. Kuntz, Jr., G.A. Petsko, Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 A, Biochemistry, 23 (1984) 2849-2857.

[16] S.V. Lepeshkevich, S.N. Gilevich, M.V. Parkhats, B.M. Dzhagarov, Molecular oxygen migration through the xenon docking sites of human hemoglobin in the R-state, Biochimica et biophysica acta, 1864 (2016) 1110-1121.

[17] M.S. Shadrina, G.H. Peslherbe, A.M. English, Quaternary-Linked Changes in Structure and Dynamics That Modulate O2 Migration within Hemoglobin's Gas Diffusion Tunnels, Biochemistry, 54 (2015) 5268-5278.

[18] J.A. Cortazzo, A.D. Lichtman, Methemoglobinemia: a review and recommendations for management, J Cardiothorac Vasc Anesth, 28 (2014) 1043-1047.

[19] C. Brown, M. Bowling, Methemoglobinemia in bronchoscopy: a case series and a review of the literature, Journal of bronchology & interventional pulmonology, 20 (2013) 241-246. 19

[20] H.A. Cooper, Methemoglobinemia caused by benzocaine topical spray, South Med J, 90 (1997) 946-948.

[21] M.F. Perutz, Preparation of haemoglobin crystals, Journal of Crystal Growth, 2 (1968) 54-56.

[22] K.L. Bren, R. Eisenberg, H.B. Gray, Discovery of the magnetic behavior of hemoglobin: A beginning of bioinorganic chemistry, Proc Natl Acad Sci U S A, 112 (2015) 13123-13127.

[23] S.A. Wilson, T. Kroll, R.A. Decreau, R.K. Hocking, M. Lundberg, B. Hedman, K.O. Hodgson, E.I. Solomon, Iron L-edge X-ray absorption spectroscopy of oxy-picket fence porphyrin: experimental insight into Fe-O2 bonding, J Am Chem Soc, 135 (2013) 1124-1136.

[24] S.A. Wilson, E. Green, Mathews, II, M. Benfatto, K.O. Hodgson, B. Hedman, R. Sarangi, X-ray absorption spectroscopic investigation of the electronic structure differences in solution and crystalline oxyhemoglobin, Proc Natl Acad Sci U S A, 110 (2013) 16333-16338.

[25] J. Yi, L.M. Thomas, G.B. Richter-Addo, Structure of human R-state aquomethemoglobin at 2.0 A resolution, Acta Crystallogr Sect F Struct Biol Cryst Commun, 67 (2011) 647-651.

[26] M.K. Safo, M.H. Ahmed, M.S. Ghatge, T. Boyiri, Hemoglobin-ligand binding: understanding Hb function and allostery on atomic level, Biochimica et biophysica acta, 1814 (2011) 797-809.

[27] D. Sehnal, R. Svobodova Varekova, K. Berka, L. Pravda, V. Navratilova, P. Banas, C.M. Ionescu, M. Otyepka, J. Koca, MOLE 2.0: advanced approach for analysis of biomacromolecular channels, J Cheminform, 5 (2013) 39.

[28] S. Prahl, Tabulated molar extinction coefficient for hemoglobin in water, Oregon Medical Laser Center, 4 (1998).

[29] A.I. Vogel, B.S. Furniss, Vogel's Textbook of practical organic chemistry, 5th ed. / ed., Longman Scientific & Technical ; Wiley, London :New York :, 1989.

[30] W. Kabsch, Xds, Acta Crystallographica Section D: Biological Crystallography, 66 (2010) 125-132.

[31] 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, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallographica Section D: Biological Crystallography, 66 (2010) 213-221.

[33] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr, 66 (2010) 486-501.

[34] Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.8, 2015.

3 THE MDM2 RING INHIBITORS INULANOLIDE A AND MA242 SHARE A CONSERVED BINDING SITE

3.1 Abstract

The murine double minute 2 (MDM2) protein is an oncogenic E3 ubiquitin ligase which shows upregulated expression in numerous cancers. MDM2 contributes to cancer cell proliferation and metastasis primarily by promoting degradation of the tumor suppressor protein p53 together with its other p53 independent functions. Previous experiments have identified compounds Inulanolide A and MA242 as dual NFAT1-MDM2 inhibitors to inhibit cancer growth. Both Inulanolide A and MA242 have been shown to bind the MDM2 RING domain, but their exact mechanism of action is still not fully understood. In this work, crystal structures of Inulanolide A and MA242 in complex with the MDM2 RING domain were determined to a resolution better than 1.7 Å. The sidechain and backbone interactions were identified in the inhibitor binding site within the MDM2 RING domain. Tyr489 is the essential residue that binds the inhibitors via hydrophobic interactions. Autoubiquitination rates of the MDM2 RING domain by an E2 ligase were modestly reduced by these inhibitors. The results suggest that shielding residue Tyr489 inhibits ubiquitination of MDM2 substrates such as p53, but still allows the MDM2 RING domain to be autoubiquitinated, leading to MDM2 degradation.

3.2 Introduction

The tumor suppressor p53 is commonly referred to as the “Guardian of the Genome” due to its role in repairing DNA damage and preventing cancer formation [1, 2]. In healthy cells that experience potentially mutagenic stress, p53 will become activated and arrest the cell cycle while also promoting the transcription and implementation of reparatory signaling cascades [3]. This activation occurs in two stages, with the first stage resulting in an increased intracellular concentration of p53 and the second stage activating p53 dependent transcription following the phosphorylation of its N-terminal domain [4]. In the event that activation levels are sustained or reach a critic threshold, p53 will induce apoptosis [5]. In order to prevent undesirable runaway signaling reactions, the intracellular and intranuclear concentration and level of p53 activation are meticulously regulated [6]. Regulatory mechanisms rely heavily upon ubiquitination signals for nuclear export, sequestration, and proteolytic degradation [7]. When the regulation or activity of p53 within the cell is disrupted, catastrophic gene mutation and uncontrolled replication can occur often leading to cancer; mutations of the *TP53* gene coding for p53 isoforms represent the single most ubiquitous mutation in cancer cells with frequency occurring at the highest rates (38-50%) in ovarian, lung, colorectal, esophageal, prostate, and breast cancers [8].

A key p53 regulatory protein found within cells is the murine double minute 2 (MDM2) E3 ubiquitin ligase, which facilitates both the shuttling of p53 out of the nucleus and its proteolytic degradation via ubiquitin transfer. Under normal conditions the MDM2 protein is maintained at a constant baseline level, but its expression is upregulated by p53 gene transcription in a negative feedback manner [9]. In many cancers the MDM2 protein is overexpressed; this results in disruption of the control mechanisms that keep an increased p53 concentration within the nucleus and cytosol [10, 11]. The MDM2 protein is comprised of four domains: an N-terminal p53 binding domain, central acidic and zinc-binding domains, and a C-terminal RING [12]. Ubiquitin transfer via MDM2 RING mediation to p53 will signal nuclear export with monoubiquitination and proteasomal degradation with polyubiquitination [13, 14].

Compounds exhibiting an inhibitory effect on p53-MDM2 interactions and ubiquitin transfer have been shown to function as tumor suppressants in cancer cell lines [15]. In previous publications, Qin *et. al.* and Wang *et. al.* identified a dimeric sesquiterpenoid natural product

Inulanolide A (InuA) found in *Inula brittanica* and compound MA242 as inhibitors of both MDM2 and the transcription factor NFAT1 (Nuclear factor of activated T-cells 1) [16,17, 18]. A strong affinity specifically for the MDM2 RING domain (A.A. 419-491) was demonstrated in pulldown assays using biotinylated InuA and MA242 [17, 18]. Additionally, MA242 was also shown to enhance the proteasome degradation of intracellular MDM2 in an NFAT1 independent manner [18]. Inhibition of NFAT1 downregulates the expression of the *mdm2* oncogene in a p53 independent manner **(Figure 3.1).** These effects were found to demonstrate profound anticancer effects *in vitro* and *in vivo* [18, 19]. This observed dual inhibition by small molecules makes them promising anticancer agents targeting the NFAT1-MDM2-p53 signaling network.

In order to understand the InuA and MA242 mechanisms of action and allow for potential structure-based drug design, more detailed knowledge about how inhibitors interact with sidechains and backbones of key residues in the binding site is essential. Previous structural studies by X-ray crystallography have provided an initial framework to experimentally examine binding interactions of the inhibitors with the MDM2 RING domain [19]. This work seeks to bridge the current gap between cellular studies and protein structures to provide a more detailed view of the mechanism of action by these potentially therapeutic compounds.

3.3 Results

3.3.1 Inulanolide A binds to an exposed hydrophobic region of the MDM2 RING C-terminus

Crystals of the MDM2 RING domain were grown using the described method. The poor solubility and bulky structure of InuA mandated long soaking times. Incorporation of InuA was only achieved by the transfer of crystals into an stock solution comprised of the crystallization well solution, 20% MPD, 10% DMSO and 10 mM InuA over a 5-day period. X-ray diffraction data were collected to a resolution of 1.48 and 1.56 Å for native and InuA soaked crystals, respectively, and a piece of extra electron density was observed in proximity to the C-terminus of a single MDM2 RING chain when the structure of the InuA soaked crystal was compared with that of the native crystal. Statistics for both diffraction data and structural refinement can be found in **[Table 3.1]**. The extra density is positioned over residue Tyr489 **[Figure 3.2]**. Binding of additional InuA molecules was not observed at the other three potential sites in the subunits within the asymmetric unit, which may be due to a lack of accessibility resulted from the tight crystal packing at the other three interfaces.

The structure of InuA and the MDM2 RING domain complex reveals that the primary binding mode is though interactions with an exposed hydrophobic region of the MDM2 RING homodimer **[Figure 3.3]**. The sidechain of Tyr489 rests approximately 3.3 Å beneath the center of the InuA structure comprised of a fused tetrahydrobenzofuran/cyclohexene-cyclohepta-furan moiety. This distance falls into the ideal range of van der Waals interactions [32]. Analogs retaining the structural moiety of cyclohexene-cyclohepta-furans also possess strong antitumor activities targeting MDM2 [17]. Sidechains of Asn447 and Phe430 and Pro431 of the other subunit in the MDM2 RING dimer contribute to hydrophobic interactions with the terminal furan in InuA. An additional hydrophobic pocket is also formed by Phe490 and Pro491 allowing for further hydrophobic interfacing with InuA. A noteworthy hydrogen bond is formed between the sidechain of His457 from the adjoined MDM2 subunit and the oxygen atom in the furan of the tetrahydrobenzofuran at a distance of 3.10 Å from N to O.

3.3.2 MA242 Shares a binding site with Inulanolide A

X-ray diffraction data were also collected to 1.68 Å resolution from crystals soaked with 10 mM MA242. Incorporation of MA242 was also facilitated using the method described for

InuA. Statistics on X-ray diffraction data and structural refinement are also found in [**Table 3.1]**. In comparison with the native structure, a piece of extra density was also observed in the same region as in the InuA-MDM2 structure, with some subtle differences resulted from differences in the molecular structures [**Figure 3.4]**. The primary binding mode is also from a central hydrophobic/π-stacking interaction between the sidechain of Tyr489 and the dihydropyrroloquinolin and chlorobenzene groups of MA242 **[Figure 3.5]**. The sidechain of Thr455 from the adjoined subunit in the MDM2 RING dimer has hydrophobic interactions with the chlorobenzene ring. On the other end of MA242, sidechains of Pro491 and Asn433 from the adjoined subunit has hydrophobic or π-NH2, respectively, interactions with the aryl component of the tosyl moiety. Hydrogen bonds are formed between the mainchain amide nitrogen of Phe490 and the sulfonamide branching from the core of MA242, as well as the non-Zn-coordinated His457 sidechain nitrogen with the dihydropyrroloquinolin keto group of MA242.

3.3.3 Autoubiquitination assays with Inulanolide A and MA242 display inhibitory effects

Autoubiquitination assays were performed on the GST-tagged MDM2 RING domain to examine the effects on ubiquitin transfer. UBA1 and UbcH5B were utilized as the E1 and E2 ligases, respectively. Fluorescein-labeled ubiquitin transfer to GST-MDM2 RING protein was observed at multiple timepoints. In the SDS-PAGE gel, the fluorescence intensity in the bands ranging between 40-100 kDa was reduced when inhibitors were included. Monoubiquitin MDM2 RING conjugates appeared at 43.5 kDa with di- and polyubiquitination as laddered bands ranging from 53 to 100 kDa **[Figure 3.6A, 3.10]**. In the presence of both InuA and MA242, autoubiquitination was reduced modestly (~32.8 and 27.8%) when compared to a DMSO control **[Figure 3.6B]**.

3.4 Discussion

The C-terminal region of the MDM2 RING domain is essential for the E3 activity of MDM2 towards p53 and other downstream targets [23]. C-terminally truncated mutants of the MDM2 RING domain demonstrate a loss of nearly all activity when examining the mediation of ubiquitin transfer to p53 [24]. Additionally, substitution of Tyr489 or Phe490 with non-aryl amino acids both demonstrated little to no E3 activity towards p53 [23], suggesting that these sidechains must play an essential role in the delivery of the ubiquitin by the E2-Ub-MDM2 ternary complex. Examination of MDM2 RING dimer structures provides some insights into the roles of these sidechains. Phe490 seems to be utilized primarily for the formation of the hydrophobic dimerization interface within the MDM2 RING domain, while Tyr489 extends outwards towards a solvent accessible region making its role in E3 activity seem significantly more nuanced **[Figure 3.7]**. In the ternary complex determined by Magnussen *et. al.* [PDB ID: 6SQR, [21]] the Tyr489 sidechain is located adjacent to the epitope of the E2-Ub with the MDM2 RING domain. Upon determining the localization of inhibitory compounds to this hydrophobic region it does not seem surprising that they act as inhibitors owing to presumable steric hinderance of ternary complex formation **[Figure 3.8 and 3.9]**. Both small molecules occupied a region in which the C-terminus of the E2-conjugated ubiquitin α-helix (Ub residues 31-36) resides spatially. In spite of these structural observations, the exact mechanism of inhibitory actions cannot be ascertained by examining the structure alone and biochemical experiments that were performed aided in the substantiation of potential effects. Mechanistically the MDM2 RING domain possesses E3 ligase activity against both target proteins and itself in the form of autoubiquitination. Potential disruption of ternary complex formation occurring in solution with the inclusion of InuA and MA242 should be observable by a reduction in ubiquitin transfer to the MDM2 RING domain. Autoubiquitination assays demonstrated an approximately 30% reduction in overall ubiquitination of GST-tagged MDM2 RING protein in the presence of inhibitors. This data combined with observed binding modes of InuA and MA242 suggests that the primary mechanism of their inhibitory activity is through disruption of ternary complex formation with coupled E2 transferases. Autoubiquitination of MDM2 has been demonstrated to enhance its E3 activity towards p53 so disruption of this mechanism is a promising target for drug development [24]. Proximity of the binding site to the S429E substitution further substantiates the observed effects as phosphorylation of Ser429 has been shown to enhance autoubiquitination by stabilization of this interface without steric disruption [21].

3.5 Methods

3.5.1 Protein Expression and Purification

The MDM2 RING domain construct (422-C-S429E/G443T-491) was cloned into the pGEX-6p-1 vector containing an N-terminal GST-tag and a mutated TEV protease cleavage sequence by PCR. Cloned fragments were verified by DNA sequencing (Eurofins Genomics). GST-MDM2 RING domain (422-C-S429E/G443T-491) was expressed in BL21 Rosetta (DE3) *E. coli* cells (Novagen). Cells were grown in 2X-YT media at 37˚C to an OD_{600nm} of 0.6 and expression induced with 0.5mM IPTG at 18°C for 16 hours. Cells were harvested following pelleting by centrifugation. Cells containing the expressed fusion protein were suspended in buffer containing 25mM TRIS-HCl, pH 7.6, 0.4M NaCl, and 1mM DTT as described by Magnussen et. al. [19]. The suspension was stirred for 1 hour on ice and then treated with 2.5mM PMSF. Cells were lysed by sonication and cell debris pelleted by centrifugation. Clarified lysate was loaded onto a 5mL GSTrap HP column (GE Healthcare). Protein was eluted using wash buffer containing 25mM TRIS-HCl, pH 7.6, 0.4M NaCl, 1mM DTT, and 10mM reduced glutathione and treated overnight with a 1:10 stoichiometric ratio of TEV protease. Cleaved protein was diluted to an NaCl concentration of 175mM using 50mM TRIS-HCl, pH 7.6 and purified by ion-exchange chromatography using a 1mL HiTrap SP HP column (Cytiva) followed by gel-filtration purification on a HiLoad Superdex 75 16/60 column (GE Healthcare) pre-equilibrated with the aforementioned lysis buffer. Fractions containing purified MDM2 RING domain were examined by SDS-PAGE to confirm purity and pooled prior to use in crystallization experiments. For autoubiquitination experiments the GST-MDM2 RING fusion protein was left uncleaved and purified solely by GSTrap and HiLoad Superdex 75 columns.

3.5.2 Crystallization

The cleaved MDM2 RING domain construct containing a retained N-terminal serine residue from the TEV protease sequence (422-Cat-S429E/G443T-491) was concentrated to 11mg/mL and crystallized by hanging drop vapor diffusion at 23°C in a 1:1 mixture with well solution containing 0.1M HEPES pH=7.5, 200mM NaCl, and 25% PEG3350. Crystal formation was observed over 1 to 3 days. After initial growth, crystals were transferred to a cryoprotectant solution containing 0.1M HEPES pH=7.5, 200mM NaCl, 25% P.E.G. 3350, 20% MPD, and 10% DMSO (control structure)/DMSO stock solution (100 mM stock, 10 mM final conc.) of InuA or MA242. Crystals were incubated in the presence of solely DMSO or DMSO/compound for 5 days prior to flash cooling in liquid nitrogen.

3.5.3 Autoubiquitination Assays

Autoubiquitination assays were carried out using a protocol derived from the work of Magnussen *et. al.* [21] with minor modifications to increase reaction times. In short, UBA1 (0.2 µM), UbcH5B (2.5 µM), and fluorescein-labeled ubiquitin (25 µM, Flr-Ub, Enzo Biosciences) were incubated for 30 minutes in buffer containing 50 mM Tris-HCl pH 7.0, 50 mM NaCl, 5 mM $MgCl_2$, and 5 mM ATP to precharge the E2 ligase with Flr-Ub. 10x GST-MDM2 RING protein solutions containing DMSO or DMSO/compound were added yielding final reaction concentrations of 0.5 µM GST-MDM2, 0.05% DMSO, and 50 µM inhibitor. Experimental timepoints were obtained by quenching reaction aliquots with SDS-PAGE loading dye containing 400mM DTT at 1 minute intervals. All reactions utilized for data acquisition were performed in triplicate. Reaction aliquots were run on 10% acrylamide SDS-PAGE gels and imaged using a BioRad ChemiDoc MP imaging system [uncropped SDS-PAGE gels are shown in **Figure 3.10**]. Fluorescence intensity of mono- and polyubiquitinated GST-MDM2 RING protein was quantified using ImageJ software [25] and plotted using Microsoft excel.

3.5.4 X-ray Data Processing and Analysis

X-ray diffraction data were collected at SER-CAT at the Advanced Photon Source, USA on the ID-22 beamline. Diffraction images were initially processed using XDS and data reduction conducted using Aimless in the data reduction CCP4i2 suite [26,27]. Molecular replacement was performed with PHENIX [28] using maximum-likelihood procedures in PHASER-MR and the coordinates from the MDM2 RING structure (PDB Code 6SQP, downloaded from the RCSB protein data bank) [21] as the search model. Structural refinement was carried out using phenix.refine. Modeling of coordinates in electron density was performed with Coot [29]. A summary of crystallographic data is shown in [**Table 3.1]**. Figure preparation was carried out using PyMOL (v 1.8, Schrodinger) [30]. 2D ligand binding figures were generated using the LigPlot+ software [31].

3.6 Figures

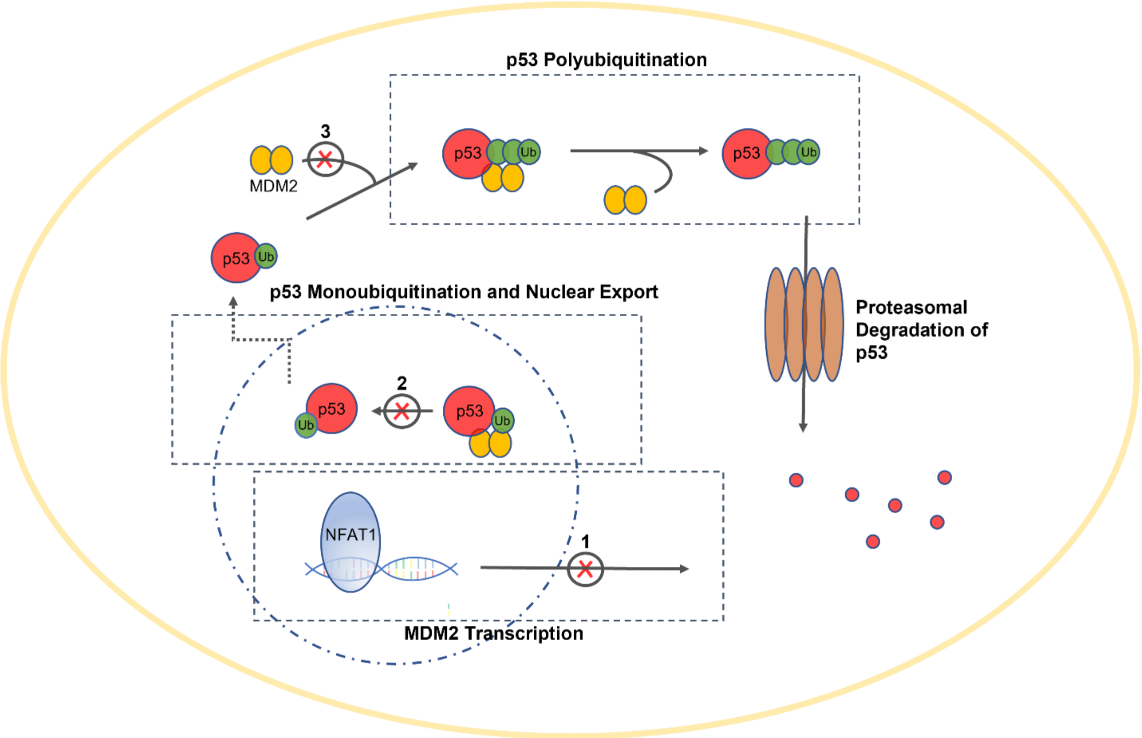


Figure 3.1 Intracellular effects of InuA and MA242

Both InuA and MA242 have demonstrated inhibitory effects on **1)** Transcription of the *mdm2* oncogene by NFAT1 **2)** Intranuclear MDM2 RING E3-mediated transfer of ubiquitin to p53 **3)** MDM2 RING E3-mediated polyubiquitination of p53 resulting in sequestration and proteolysis. Inhibition of MDM2 within the nucleus (**2**) and the cytosol (**3**) may prevent the disruption of cellular functions of p53

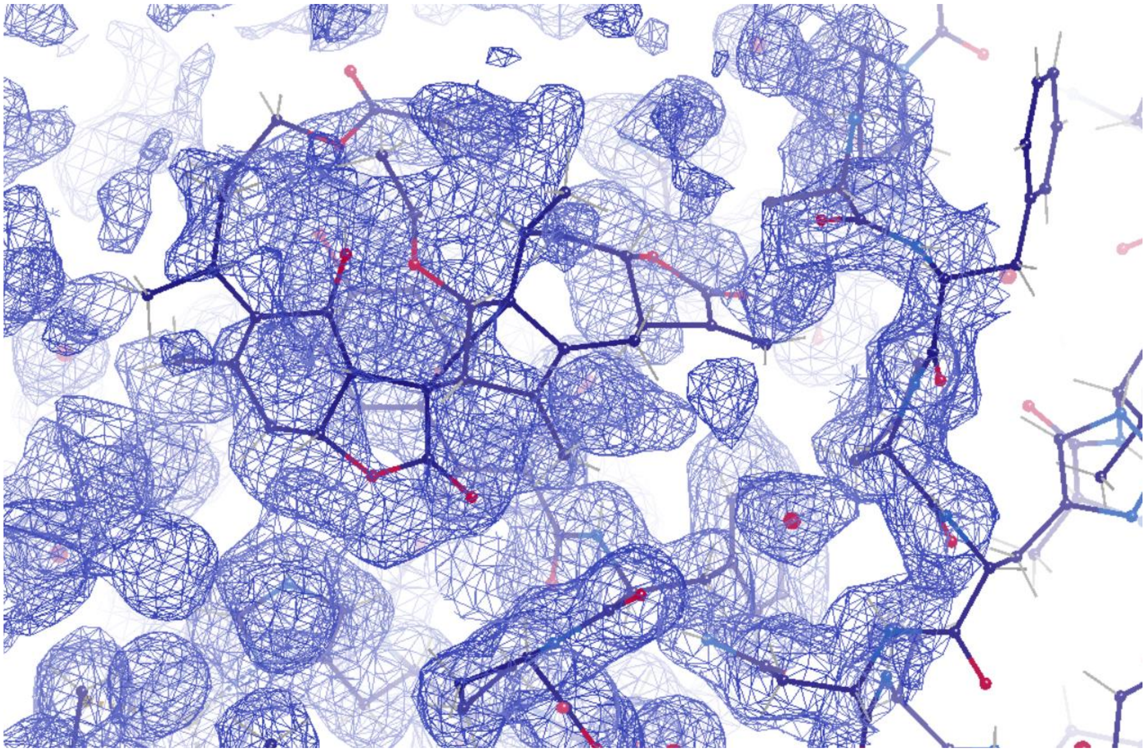


Figure 3.2 2mF0-DFc map of the inulanolide A soaked structure at 0.5σ
A top down view of the InuA binding site located ~3 Å above Tyr489 of chain A in homodimer 1 of the asymmetric unit.

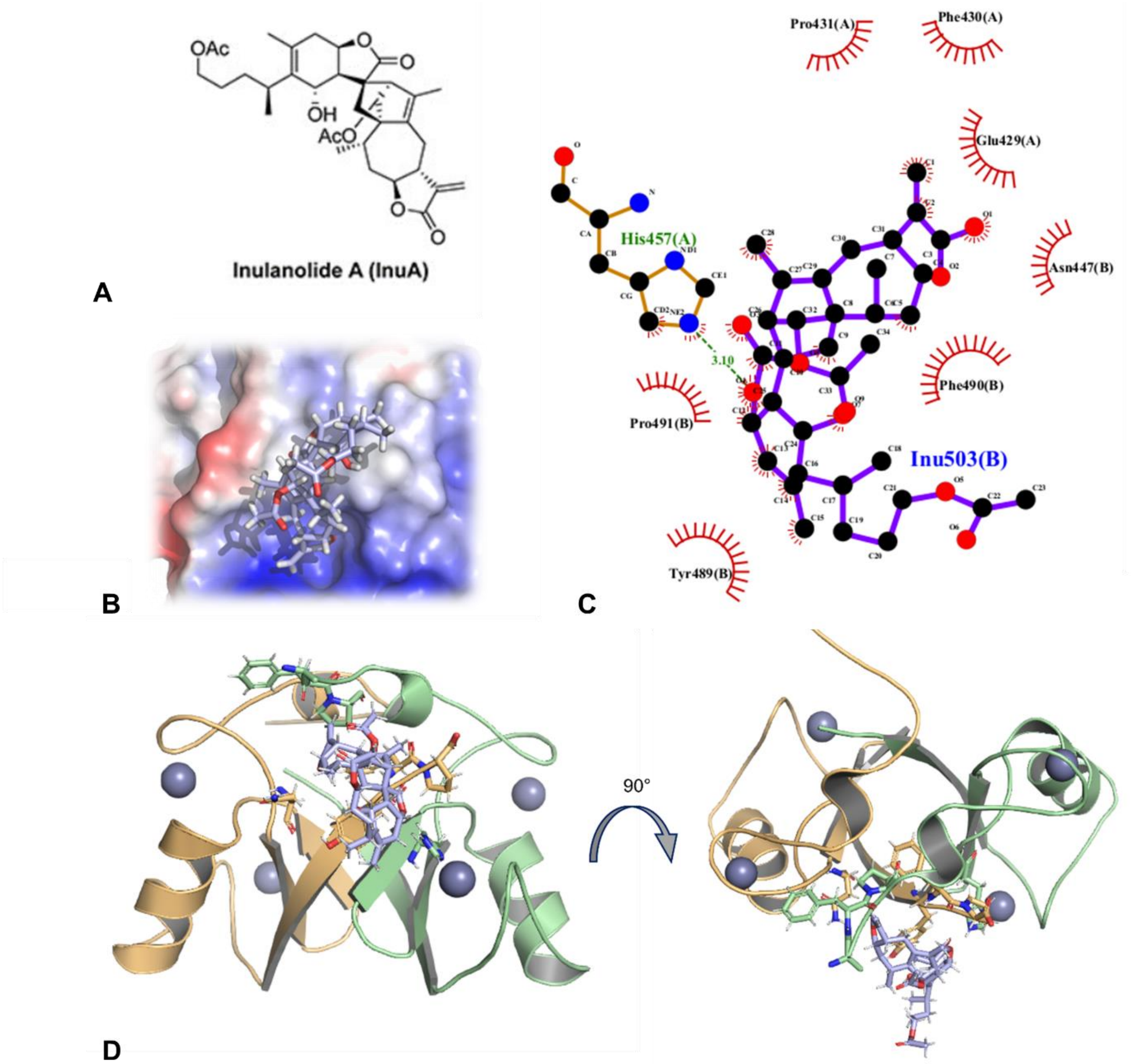


Figure 3.3 Interactions of InuA with the MDM2 RING homodimer
A) The molecular structure of inulanolide A **B**) Localized surface electrostatics of the homodimer structure. Inulanolide A is shown in ball and stick and curves along a distinguished solvent-exposed hydrophobic stretch **C**) Ligplot+ 2D binding diagram displaying key amino acid contacts comprising the binding site of the small molecule. Hydrogen bonding with the imidazole nitrogen of His457 from chain B of the homodimer is observed at a distance of 3.1 Å. **D**) InuA localization at the interface of the MDM2 RING (InuA (light blue) homodimer chain A (green) and chain B (orange)). Residue Tyr489 is labeled.

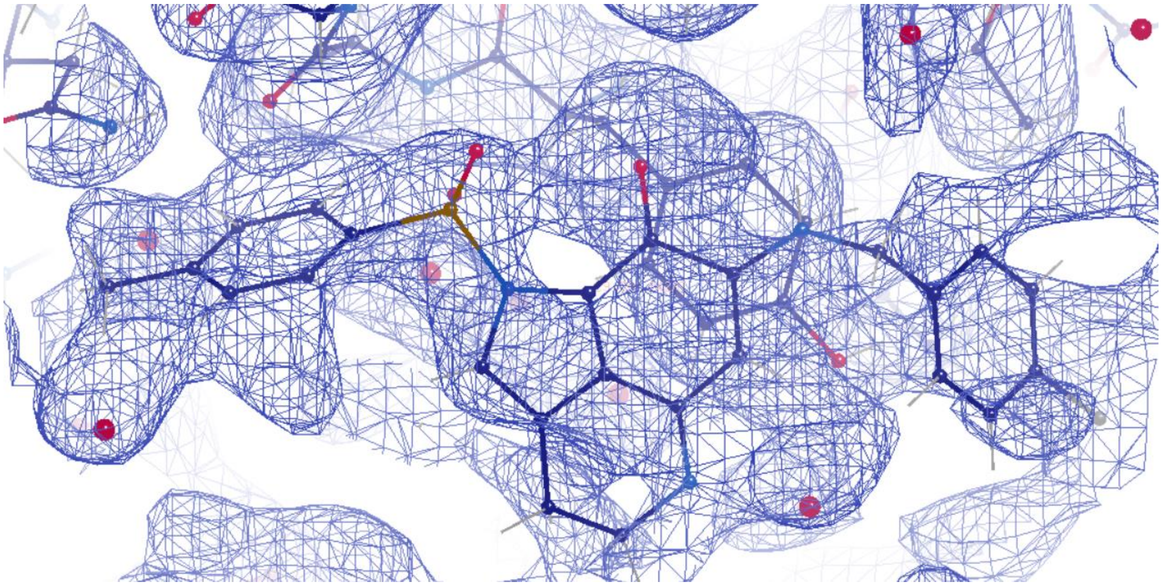


Figure 3.4 2mF0-DFc map of the MA242 soaked structure at 0.5σ
A top down view of the MA242 binding site located 3.4 Å above Tyr489 of chain A in homodimer 1 of the asymmetric unit.

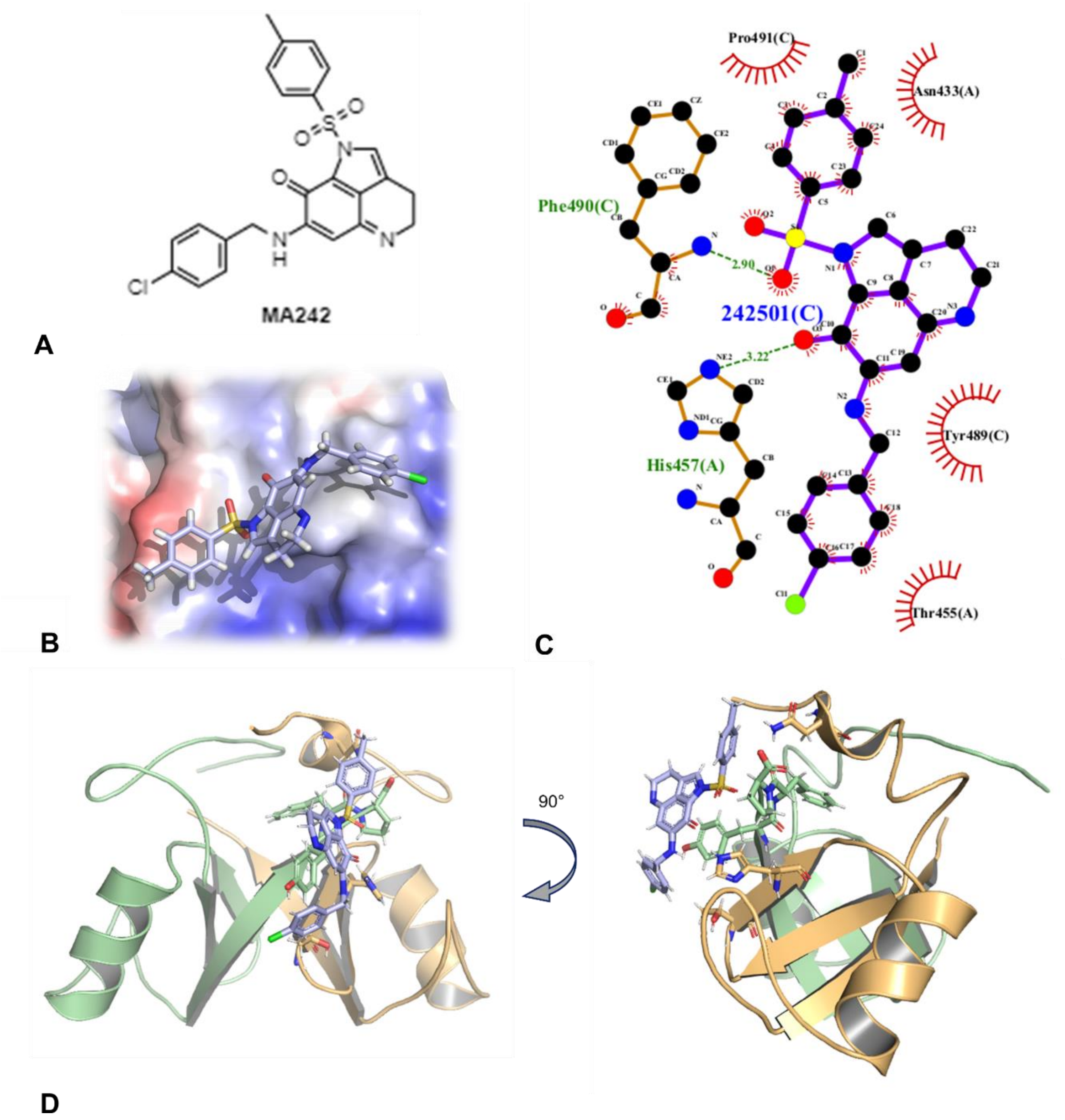


Figure 3.5 Interactions of MA242 with the MDM2 RING homodimer
A) The molecular structure of MA242 **B**) Localized surface electrostatics of the homodimer structure. MA242 is shown in ball and stick and also curves along the exposed C-terminal hydrophobic stretch **C**) Ligplot+ 2D binding diagram displaying key amino acid contacts comprising the binding pocket of the small molecule. Hydrogen bonding with the imidazole nitrogen of His457 from chain A and the backbone amide nitrogen of Phe490 from chain C of the homodimer are observed at a distance of 3.22 and 2.99 Å respectively. Additional contacts occurring are primarily hydrophobic interactions. **D**) MA242 localization at the interface of the MDM2 RING (homodimer chain A (green) and chain C (orange)). Residue Tyr489 is labeled.

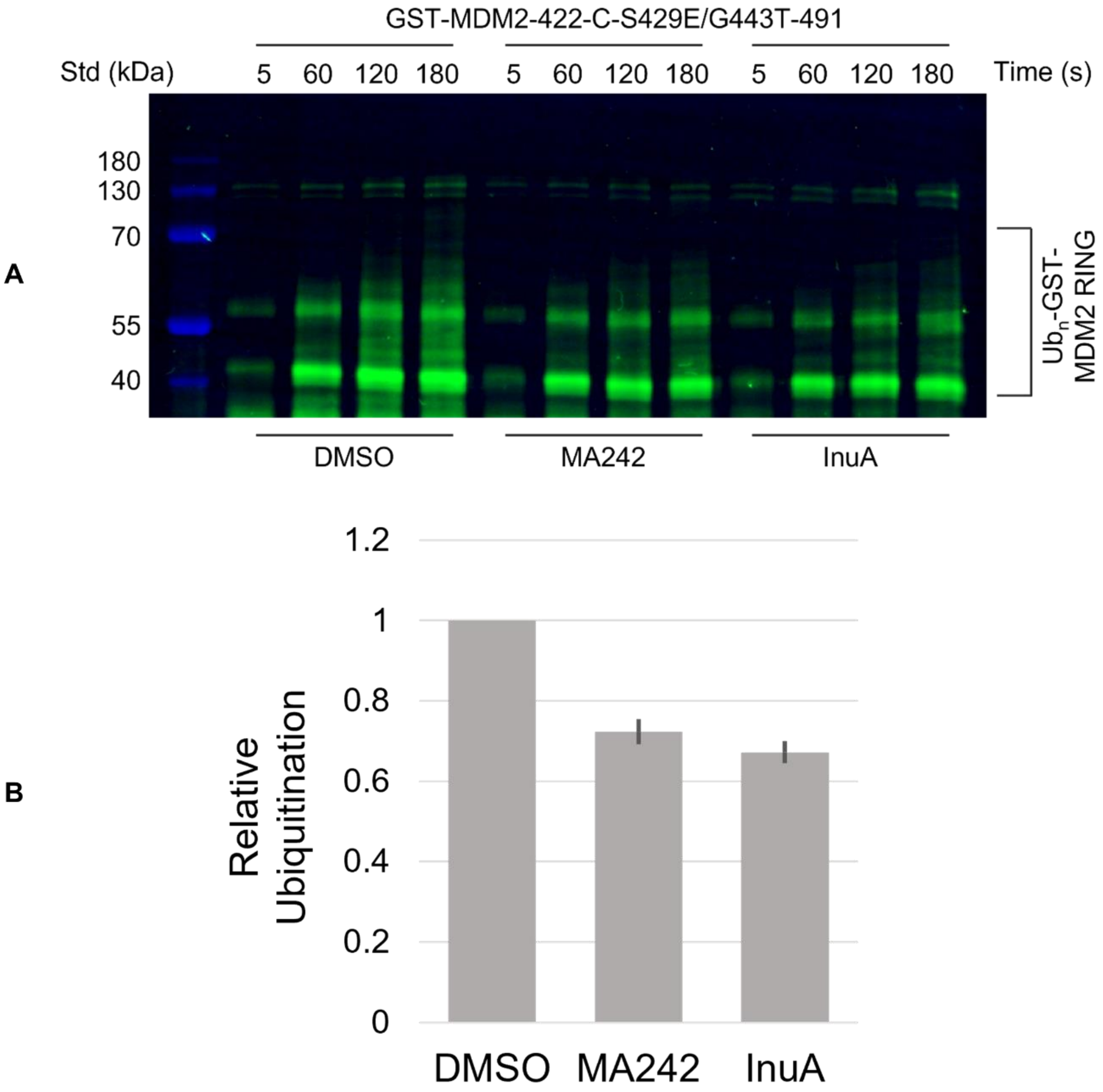


Figure 3.6 Effects on autoubiquitination activity in the presence of MA242 and InuA
A) Reduced SDS-PAGE showing auto-Flr-ubiquitination of GST-MDM2-422-C-S429E/G443T-491 by UbcH5B E2 ligase in the presence of DMSO (control) and inhibitor compounds at 50 µM. Monoubiquitination bands appear at 43.5 kDa with di- and polyubiquitination appearing directly above as laddered bands. Fluorescein-labeled ubiquitin conjugation increases the substrate molecular weight by an additional 9.5 kDa per conjugation. **B)** Relative levels of ubiquitination following peak integration at 3 minutes reaction time. Reactions were performed in triplicate and uncropped gels are shown in **Figure 3.10**.

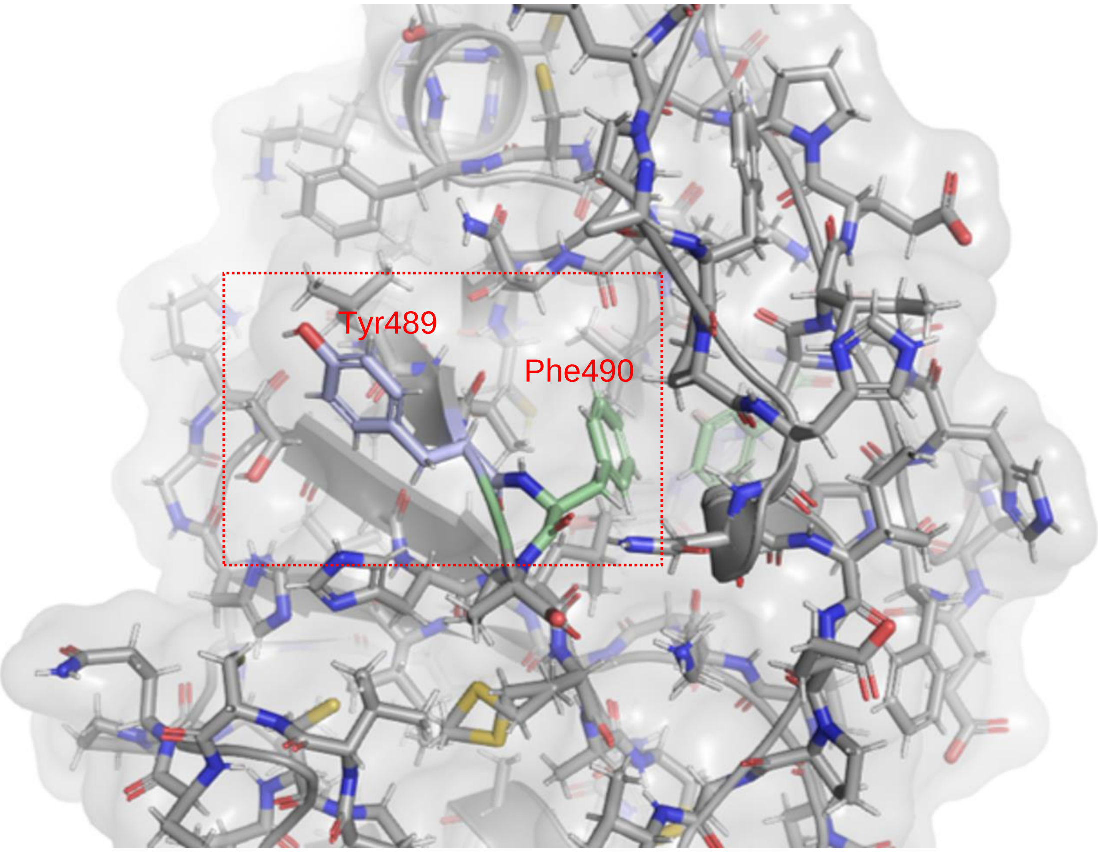


Figure 3.7 Intramolecular interactions of Tyr489 and Phe490 within the MDM2 RING Homodimer
Tyrosine 489 (light blue) extends outwards into the solvent accessible area on the exterior of the MDM2 RING homodimer with orientation directed by hydrophobic interaction with an adjacent isoleucine residue. Phenylalanine 490 (light green) extends inwards and interacts with the mirrored phenylalanine residue of the conjoined MDM2 RING monomer in a large hydrophobic pocket comprising the core of the dimeric interfacial structure.

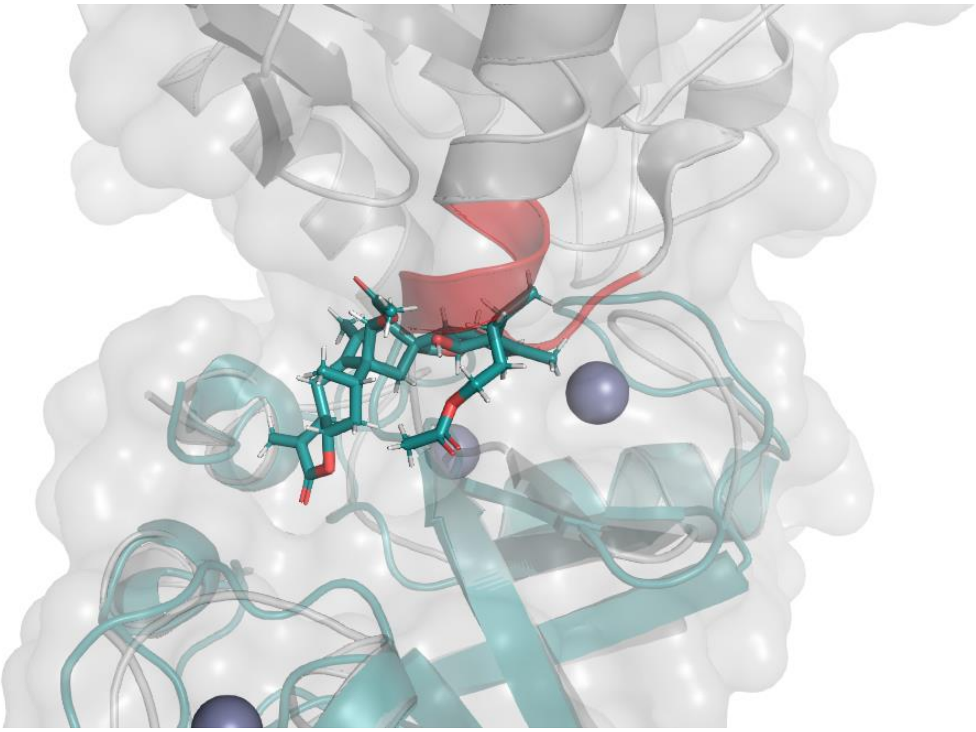


Figure 3.8 Steric disruption of UbcH5B-Ub with MDM2 RING Domain by InuA
InuA-bound MDM2 RING domain (teal) alignment with the previously determined UbcH5B-Ub-MDM2 RING structure (grey, 6SQR, [21]). The C-terminus of ubiquitin helix 1 (residues 31-36, red) occupies the same region as the determined binding site of InuA.

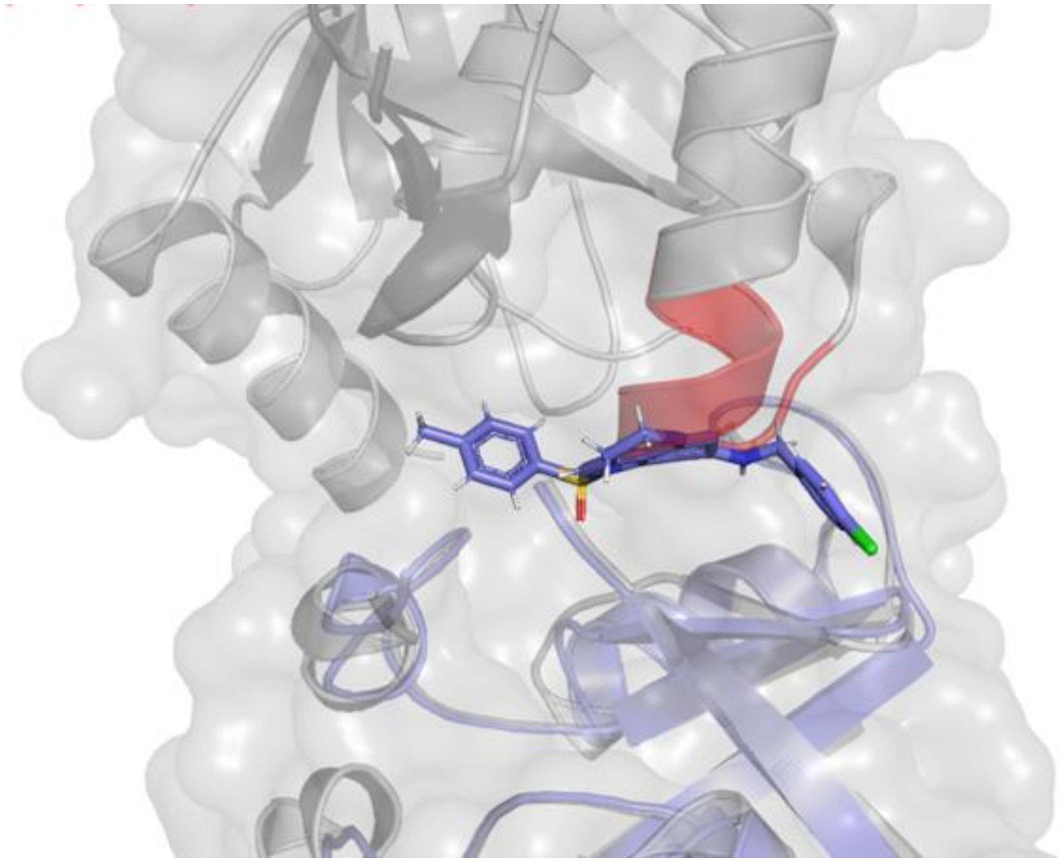


Figure 3.9 Steric disruption of UbcH5B-Ub with MDM2 RING Domain by MA242
MA242-bound MDM2 RING domain (slate) alignment with the determined UbcH5B-Ub-MDM2 RING structure (grey, 6SQR [21]). The C-terminus of ubiquitin helix 1 (residues 30-36, red) occupies the same coordinates as the determined binding site of MA242. Presumable steric disruption of ternary complex formation is observed.

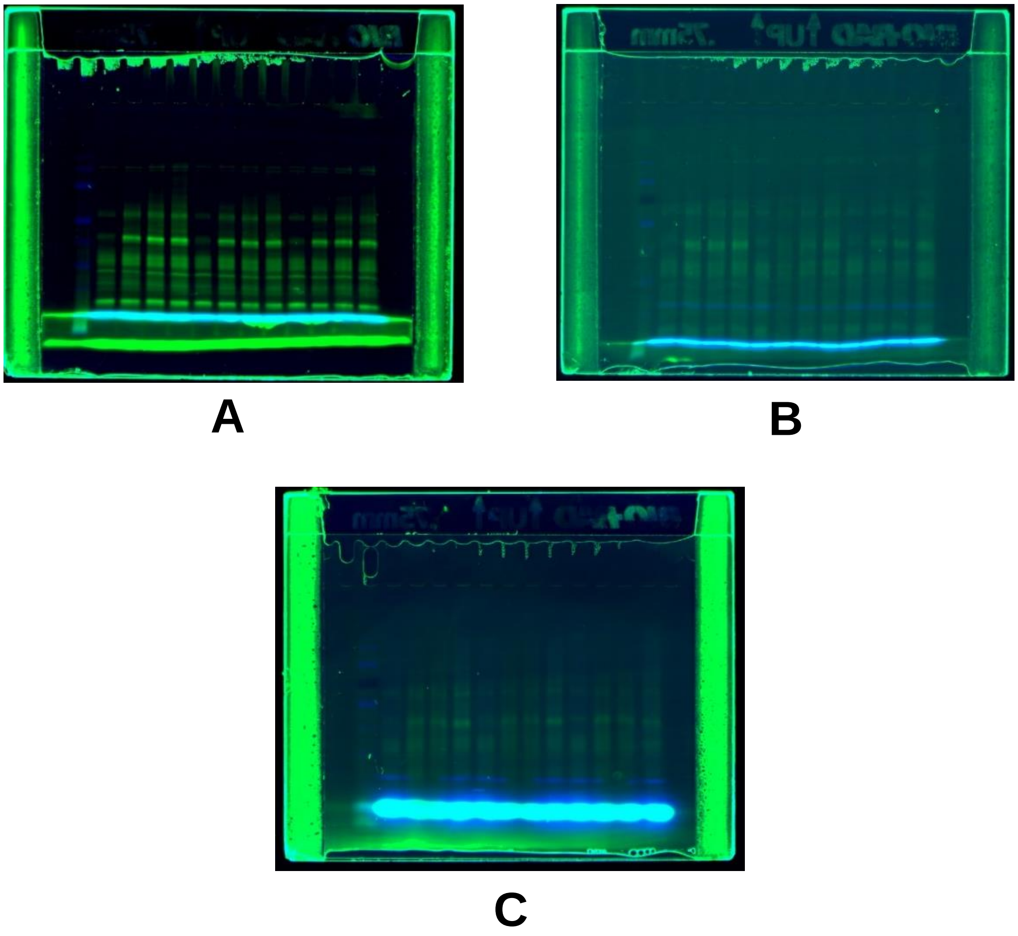


Figure 3.10 Uncropped and unsplit autoubiquitination assay gels
Reduced SDS-PAGE gels of autoquitination trials 1-3 (A, B, C). Lanes left-to-right: ladder, lanes 1-4 are DMSO control at 5, 60, 120, and 180 second time points, 5-8 and 9-12 are identical reactions containing 50 µM MA242 and 50 µM InuA respectively.

Table 3.1 X-ray Diffraction and Refinement Statistics for MDM2 RING structures

	MDM2 422-C-S429E/G443T-491 Native	MDM2 422-C-S429E/G443T-491 MA242	MDM2 422-C-S429E/G443T-491 InuA
Space Group	$P1\ 2_1\ 1$	P1	$P1\ 2_1\ 1$
Cell Dimensions			
a, b, c (A)	29.29, 39.832, 103.171	29.351, 39.794, 52.095	29.145, 39.618, 102.656
α,β,γ (°)	90, 94.019, 90	84.438, 85.878, 89.981	90, 94.111, 90
Resolution (A)	31.44-1.48 (1.53-1.48)	30.07-1.68 (1.74-1.68)	31.33-1.56 (1.62-1.56)
R_{merge} (%)	8.313 (39.91)	7.392 (38.5)	5.587 (29.33)
I/σI	12.74 (3.89)	8.53 (2.40)	19.39 (4.85)
Completeness (%)	98.25 (99.02)	96.27 (94.82)	99.68 (99.67)
Multiplicity	4.9 (5.0)	2.7 (2.7)	3.6 (3.6)
CC (1/2)	0.998 (0.942)	0.996 (0.861)	0.999 (0.936)
Wilson B (A^2)	11.82	19.14	11.23
Refinement			
Resolution	31.44-1.48	30.07-1.68	31.33-1.56
No. of reflections	191283 (19875)	68936 (6947)	122160 (12127)
R_{work}/R_{free}	0.1620/0.1903	0.1577/0.1885	0.1525/0.1852
No. of atoms	2275	2228	2326
Protein	2031	2026	2052
Ligand/ion	12	60	98
Water	232	162	223
B factors			
Protein	18.19	28.49	17.14
Ligand/ion	15.27	68.4	59.46
Water	27.41	32.04	24.51
R.m.s. deviations			
Bond lengths (A)	0.009	0.012	0.01
Bond angles (°)	1.02	1.18	1.05
Ramachandran			
Favored (%)	95.22	95.2	94.14
Outlier (%)	0	0	0

3.7 References

[1] Toufektchan E, Toledo F. The Guardian of the Genome Revisited: p53 Downregulates Genes Required for Telomere Maintenance, DNA Repair, and Centromere Structure. Cancers (Basel). 2018;10(5).

[2] Surget S, Khoury MP, Bourdon JC. Uncovering the role of p53 splice variants in human malignancy: a clinical perspective. Onco Targets Ther. 2013;7:57-68.

[3] Kastenhuber ER, Lowe SW. Putting p53 in Context. Cell. 2017;170(6):1062-78.

[4] Kruse JP, Gu W. Modes of p53 regulation. Cell. 2009;137(4):609-22.

[5] Nag S, Qin J, Srivenugopal KS, Wang M, Zhang R. The MDM2-p53 pathway revisited. J Biomed Res. 2013;27(4):254-71.

[6] Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science. 2004;303(5659):844-8.

[7] Pant V, Lozano G. Limiting the power of p53 through the ubiquitin proteasome pathway. Genes Dev. 2014;28(16):1739-51.

[8] Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2(1):a001008.

[9] Huun J, Gansmo LB, Mannsaker B, Iversen GT, Sommerfelt-Pettersen J, Ovrebo JI, et al. The Functional Roles of the MDM2 Splice Variants P2-MDM2-10 and MDM2-5 in Breast Cancer Cells. Transl Oncol. 2017;10(5):806-17.

[10] van den Broek AJ, Broeks A, Horlings HM, Canisius SV, Braaf LM, Langerod A, et al. Association of the germline TP53 R72P and MDM2 SNP309 variants with breast cancer survival in specific breast tumor subgroups. Breast Cancer Res Treat. 2011;130(2):599-608.

[11] Boersma BJ, Howe TM, Goodman JE, Yfantis HG, Lee DH, Chanock SJ, et al. Association of breast cancer outcome with status of p53 and MDM2 SNP309. J Natl Cancer Inst. 2006;98(13):911-9.

[12] Kostic M, Matt T, Martinez-Yamout MA, Dyson HJ, Wright PE. Solution structure of the Hdm2 C2H2C4 RING, a domain critical for ubiquitination of p53. J Mol Biol. 2006;363(2):433-50.

[13] Kubbutat MH, Jones SN, Vousden KH. Regulation of p53 stability by Mdm2. Nature. 1997;387(6630):299-303.

[14] Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. Nature. 1997;387(6630):296-9.

[15] Qi SM, Cheng G, Cheng XD, Xu Z, Xu B, Zhang WD, et al. Targeting USP7-Mediated Deubiquitination of MDM2/MDMX-p53 Pathway for Cancer Therapy: Are We There Yet? Front Cell Dev Biol. 2020;8:233.

[16] Zhang X, Zhang Z, Cheng J, Li M, Wang W, Xu W, et al. Transcription factor NFAT1 activates the mdm2 oncogene independent of p53. J Biol Chem. 2012;287(36):30468-76.

[17] Qin JJ, Wang W, Voruganti S, Wang H, Zhang WD, Zhang R. Identification of a new class of natural product MDM2 inhibitor: In vitro and in vivo anti-breast cancer activities and target validation. Oncotarget. 2015;6(5):2623-40.

[18] Wang W, Cheng J-W, Qin J-J, Hu B, Li X, Nijampatnam B, Velu SE, Fan J, Yang X-R, Zhang R. MDM2-NFAT1 Dual Inhibitor, MA242: Effective against hepatocellular carcinoma, independent of p53. Cancer Lett. 2019;459:156-167

[19] Magnussen HM, Huang DT. Identification of a Catalytic Active but Non-Aggregating MDM2 RING Domain Variant. J Mol Biol. 2021;433(5):166807

[20] Qin JJ, Li X, Wang W, Zi X, Zhang R. Targeting the NFAT1-MDM2-MDMX Network Inhibits the Proliferation and Invasion of Prostate Cancer Cells, Independent of p53 and Androgen. Front Pharmacol. 2017;8:917.

[21] Magnussen, H.M., Ahmed, S.F., Sibbet, G.J. et al. Structural basis for DNA damage-induced phosphoregulation of MDM2 RING domain. Nat Commun 11, 2094 (2020).

[22] Uldrijan S, Pannekoek WJ, Vousden KH. An essential function of the extreme C-terminus of MDM2 can be provided by MDMX. EMBO J. 2007;26(1):102-12.

[23] Poyurovsky, M. V., Priest, C., Kentsis, A., Borden, K. L., Pan, Z. Q., Pavletich, N., & Prives, C. (2007). The Mdm2 RING domain C-terminus is required for supramolecular assembly and ubiquitin ligase activity. *The EMBO journal*, *26*(1), 90–101.

[24] Ranaweera RS, Yang X. Auto-ubiquitination of Mdm2 enhances its substrate ubiquitin ligase activity. J Biol Chem. 2013;288(26):18939-46.

[25] Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 2012. *9*(7), 671–675.

[26] Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66:125-32.

[27] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67:235-42.

[28] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66:213-21.

[29] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr D Biol Crystallogr, 66 (2010) 486-501.

[30] Schrodinger, LLC, The PyMOL Molecular Graphics System, Version 1.8, 2015.

[31] Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 1995;8(2):127-34

[32] Israelachvili, J., & Pashley, R. (1982). The hydrophobic interaction is long range, decaying exponentially with distance. *Nature*, *300* (5890), 341–342

4 STRUCTURAL INSIGHTS INTO THE DYNAMIC REGULATORY MECHANISMS OF THE TRANSCRIPTION FACTOR PU.1

4.1 Abstract

Transcription factors containing an ETS-family DNA binding domain typically recognize a core consensus containing a canonical 5’-GGA(A/T)-3’ sequence motif. Divergence of ETS-family member localization within the genome is observed with variability in flanking sequences through mechanisms which are not fully understood. An essential ETS-family transcription factor, PU.1, demonstrates these classical characteristics and represents a useful model for probing the effects of flanking sequence variability on binding tunability. Herein we examine the structural observations from altered flanking sequences at a dramatically improved resolution from previously determined structures.

4.2 Introduction

Transcription factors constitute a significant functional component of the eukaryotic proteome. Present in virtually all human tissue types, transcription factors comprise ~6% of expressed genes [1]. As DNA-binding proteins, their DNA sequence selectivity, displayed as “motifs,” represents core attributes in our knowledge base of transcription factors [2]. The binding motifs of transcription factors, for which comprehensive catalogs have been curated, are indispensable for the bioinformatic assignment of non-protein-coding sequences. At the same time, motif matching with factor-enriched sequences in ChIP-seq data [3] establishes the general importance of intrinsic target specificity in transcription factor function *in vivo* [4]. The structural origins of the DNA sequence preferences of transcription factors [5], as primarily embodied by their DNA-binding domains (DBDs), therefore remain fundamental to our understanding of gene regulation.

A long-standing challenge in transcription factor/DNA recognition is posed by the selectivity mechanisms of proteins harboring DBDs that are highly structurally homologous. The ETS family of transcription factors, of which 28 members are expressed in every major tissue of humans, has been an important model for this problem [6-8]. All ETS proteins share a winged helix-turn-helix (wHTH) DBD of ~85 residues, known as the ETS domain [9]. At a gross level, ETS domains are tightly conserved in structure. In site-specific complexes, the ETS domain inserts a recognition helix (H3) into the DNA major groove harboring the core consensus 5’-GGA(A/T)-3’ while making additional backbone-mediated contacts with sequences flanking both ends of the consensus. The variations in the flanking sequences of ETS-binding motifs are important because they serve as the basis of categorization of ETS family members into four classes, I to IV [7].

To date, the mechanisms of flanking sequence discrimination by ETS proteins remain enigmatic. In many structures of ETS/DNA complexes, contacts along the interface are markedly segmented, where flanking bases in many structures are bound via backbone phosphate or sugar contacts. This feature has suggested an indirect readout mechanism in which the DNA senses the sequence-dependent propensity of DNA conformation [10]. While this interpretation dovetails with the currently favored emphasis in DNA shape [11], structural definition of this propensity remains elusive. Moreover, several co-crystallographic (4MHG, of ETV6) and solution NMR structures (2STW, of Ets-1) of ETS members show evidence of contact with flanking nucleobases. These examples of direct readout at flanking positions involve residues near the transition of H3 and a preceding loop (an extended turn in wHTH). In the NMR structures, the sidechains of these residues are notably dynamic, exhibiting various conformations that alternately contact a 5’-flanking nucleobase or avoid DNA altogether. These features, together with the absence of a viable indirect readout mechanism, prompted us to ask whether dynamic direct readout might serve as the mechanism of flanking sequence selection. To this end, an attractive model is PU.1/Spi-1, a Class III ETS-family member that exhibits distinct preferences for 5’-flanking sequences from the vast majority of its ETS relatives found in Classes I and II.

PU.1 was the first ETS domain to be co-crystallized with a sequence-specific DNA [12]. While a landmark structure [13] (1PUE), the protein carried a Gln → Glu point mutation at murine residue 228 (human residue 226) in the recognition helix, precisely the same neighborhood where direct readout of the 5’-flanking region has been implicated. PU.1 and the other class III ETS members (Spi-B and Spi-C) are indispensable regulators of hematopoiesis and immunity [14-17] and are the targets of molecular lesions in an array of hematologic malignancies [18-22]. Since this Gln residue represents a class III-restricted evolutionary difference in the recognition helix, which is the most conserved primary structure among ETS proteins throughout the metazoan [23], the unfortunate mutation in 1PUE risked obscuring our understanding of site selection by this biologically significantly sub-family. To resolve the role of this Gln in flanking sequence selection, we therefore re-examined the DNA-bound complex. We co-crystallized the sequence-corrected ETS domain of human PU.1 (which is identical to the murine ortholog) with high-affinity and low-affinity DNA targets containing the 5’-GGA(A/T)-3’ sequence [24], with a view of maximizing the potential for capturing direct readout. These new co-crystallographic structures of PU.1/DNA, which were both solved to a resolution of 1.28 Å, unveiled how PU.1 recognized a highly preferred flanking sequence by direct readout via its native loop-H3 Gln residue (human position 226). The dynamic nature of this selection was evaluated at high resolution by molecular dynamics (MD) simulations of models templated to the new co-crystal structure.

4.3 Results

4.3.1 Characteristics of the New High-Affinity PU.1/DNA Co-crystal Structure

The high-affinity DNA complex of human ETS domain of PU.1 (HA) yielded crystals that diffracted to a resolution of 1.28 Å [**Table 4.1**]. With a Wilson B-factor of 17.4 Å², the present co-crystal was significantly more ordered and better resolved than the previous structure (1PUE) at 2.10 Å and a Wilson B-factor of 26.0 Å². The high resolution of the data permitted the mapping of 788 out of 1,117 nominal hydrogens during crystallographic refinement [**Figure 4.1A**]. Comparison of the crystallographic B-factors showed significantly narrower distributions for our structure relative to 1PUE, with similar median B-factors and spatial distributions for both structures [**Figures 4.1B and 4.6**].

Although we had designed the DNA sequence with identical length and two-nucleotide Watson-Crick (WC) overhangs as reported for 1PUE [12], the DNA termini in the determined structures formed crystal contacts with adjacent DNA as short triplexes instead [**Figure 4.1C**]. Each triplex consisted of a triple stack of WC pairs from one complex interacting in the major groove with a third base from the other complex. The most distal base triplet exhibited canonical Hoogsteen base pairing of the AT*T type. The terminal protruding nucleotide (dT32) was displaced and crystallographically unresolved. Importantly, positions contacted by protein remained well-behaved as evidenced by canonical WC base-pairing and characteristic glycosidic torsion angles for B-form DNA duplexes in single crystals [**Table 2.2**]. In addition, upon alignment of the protein with its counterparts in 1PUE, the residual RMSD between backbone DNA atoms at corresponding protein-contacted base positions (which were different sequences) were only 0.8

and 1.1 Å, respectively [**Figure 4.6A**]. This close alignment indicated that the PU.1-bound DNA site was not significantly perturbed by the terminal triplexes.

To further establish that our structure represented an improved co-crystallographic model of the high-affinity PU.1 ETS/DNA complex, we compared the structures of the DNA-bound proteins. Following crystallographic refinement, the backbone heavy atoms of the high-affinity structure and the two copies in the 1PUE asymmetric units aligned to 0.26 Å. Over 90% of the ψ and ϕ backbone dihedral angles were within 10% in value, reflecting a structural core that was robust to crystallization conditions [**Figure 4.6B**]. We note that the asymmetric unit in 1PUE consisted of two conformationally non-equivalent complexes. Residues in our structure which lacked agreement with both 1PUE complexes were found only in the first and last few terminal residues (Figure 1C). Notably, the C-terminal end of the resolved primary structure exhibited well-defined albeit short helical structure (residues 254 to 258), which was not recognized as such in 1PUE. This short helix H4 was also observed in the unbound NMR solution structure of the murine PU.1 ETS domain (5W3G), supporting its postulated existence in the canonical motif of ETS domains in general [9], and marking the transition point for the C-terminal intrinsic disordered region.

Turning to the sidechains, we compared the sidechain dihedral angles of our structure and those of 1PUE [**Figure 4.7A**]. Residues exhibiting the most divergent conformations are found near crystal contacts, which were different between the two co-crystals. In our crystals, contacts were primarily mediated by a patch on the protein surface formed by helices H1, H2, and the short intervening elements. A minor crystal contact is made by the C-terminus of the recognition helix H3. [**Figure 4.7B**]. An interesting feature was observed in H2 at residue R212, which adopted two alternative conformations at approximately equal (45%:55%) occupancies. Both conformations were engaged in a salt bridge with D184 (1.7 Å and 2.5 Å), suggesting two presumably iso-energetic states. In 1PUE, different packing in the co-crystal resulted in the two corresponding residues being spaced 5 Å apart. In addition to crystal packing, several “wing” residues (around S3/S4) contacting core-flanking DNA bases, which differ from 1PUE, exhibit variations in sidechain conformations between the two models.

To complete our validation, we compared the hydration patterns in the high-affinity and 1PUE structures in order to determine the conservation of ordered water molecules in the two co-crystals. We identified, following alignment of the DNA-bound proteins, overlapping water molecules among the high-affinity structure and both non-equivalent complexes in 1PUE [**Figure 1D**]. At a cutoff deviation of 1.4 Å (the radius of a water molecule) among the three models, we identified 16 overlapped waters, 14 of which share hydrogen bond-competent geometry with both (i.e., bridging) protein and DNA. More stringent cutoffs were met at 0.7 Å (11 water) and 0.35 Å (2 water) by these bridging water molecules. Considering the different crystallization conditions and packing between the two co-crystals, we conclude that these interfacial water molecules reflected the intrinsic hydration preferences of PU.1 and represented the most stable members of the thermodynamic ensemble that is detectable in solution by osmotic pressure experiments [26, 37-39].

4.3.2 An Evolutionarily Divergent Glutamine Residue Integrates the Recognition of Core and Flanking DNA Positions via Alternate Conformations

Having established our high-affinity complex structure as a valid comparator of sequence-corrected PU.1 to the existing structure, we proceeded to interrogate the target of our structure, Q226, which was mutated to its ancestral Glu residue in 1PUE. This mutation represented the only difference between the two structures in the crystallographically resolved region. The 2Fo-Fc map at 1.0 σ in the neighborhood around Q226 showed two distinct densities in addition to that the continuous density of the sidechain [**Figure 2A**]. The shortest distances from the additional densities to Cγ and Cδ of Q226 were only 1.9 and 2.3 Å, respectively, below the sum of the vdW radii for a pair of non-bonded second-row elements. To avoid clashes, these distances must indicate mutually exclusive occupancies relative to the sidechain density and strongly suggested an alternate conformation of the Q226 sidechain. In support of this hypothesis, increasing stringency of the 2Fo-Fc map revealed progressive attenuation in density along the sidechain. We considered the possibility that the extra densities represented other components, such as water or acetate ions from the crystallization. However, the distance between the two excess densities was 2.2 Å, which was shorter than the H-bonding range for two water oxygen atoms or the expected oxygen-oxygen distance in ground-state acetate (2.4 Å) [**Figure 4.8**]. Furthermore, the unequal magnitudes of the electron densities are not consistent with an unrestrained small molecule (Figure 2A). Based on this analysis, we fitted the densities to two occupancies for the Q226 sidechain at 61% and 39% [**Figure 2B**]. The former conformation was similar to the mutated Glu residue in 1PUE, pointing down into the DNA major groove and contacting O6 of G27 flanking the core consensus (on the 5'-CCTT-3' strand) via a bridging water molecule. (In the second complex in 1PUE, the Glu residue makes the same contact minus the water in a slight variation of conformation.) In contrast, the latter "up" conformation directly contacted G6 (N7) and C7 (exocyclic NH_2), both of which are flanking positions, as well as G8 (O6) via a bridging water in the core consensus. G8 is the target of direct readout by R233, a signature core-contacting residue of the ETS domain. Thus, the "up" conformation of Q226 directly couples recognition of the flanking sequences with the 5’-GGAA-3’ consensus, a feature that is not present in 1PUE or co-crystal structures of other ETS members.

The apparently unique interactions mediated by the Class III-restricted Gln residue prompted us to interrogate it relative to Glu which is found in the other classes. To do so, we performed explicit-solvent molecular dynamics (MD) simulations to address the dynamics of Q226 and the effect of its mutation to Glu (Q226E), using our high-affinity structure as a template. Backbone and sidechain dynamics were evaluated separately from triplicate equilibrated trajectories **[Figure 4.9]** of the two complexes. Backbone dynamics in terms of per-residue RMS fluctuation (RMSF) showed negligible relative differences between the two complexes [**Figure 3A**]. Position 226 scored at the bottom quartile of backbone RMSF^2 regardless of residue identity. For the sidechains, the Q226E mutant complexed exhibited distinctive differences in per-residue RMSF. Excluding the terminal residues, sidechain RMSF differences were not correlated with altered solvent exposure, with the notable exception at the mutated position 226. The wildtype Gln residue at this position was significantly more excluded from solvent while exhibiting increased conformational dynamics [**Figure 3B**]. To probe the sidechain dynamics at position 226 in terms of DNA interactions, we determined the minimum separation distances between the sidechain with DNA. The time trajectories clearly showed a significantly broader ensemble of DNA contacts by the wildtype residue relative to E226 [**Figure 3C**]. Closer scrutiny revealed that the closer contacts by Q226 were made with 5’ flanking bases than the E226 mutant [**Figure 3D**].

4.3.3 Structural Variability Between High- and Low-Affinity DNA complexes

The structure of PU.1 DNA-binding domain in complex with the described low-affinity DNA sequence was also solved to 1.28 Å resolution **[Table 4.1]** and displayed identical packing and morphologic characteristics [**Figure 4.10A]**. Alignment with our high-affinity structure demonstrated identical protein backbone positions. Examination of sidechain occupancy at the Q226 position demonstrated a loss of alternate rotameric density, and B-factor distribution of the phosphate backbone in the coupled strand displayed a decrease in relative B-factors when compared to those of the high affinity structure suggesting a decrease in phosphate backbone motility **[Figure 4.11]**. Comparison of the spatial orientation of the dA6 N7 position of the flanking purine residue through which Q226 hydrogen bonds in the high-affinity structure demonstrates only a 0.1 Å outward shift from the Q226 β carbon. A larger 0.8 Å shift in the amine position of dC7 to dA7 between the two structures may aid in stabilization of the keto group in the alternate conformer as a 0.6 Å increase in O – N hydrogen bonding distance would be observed **[Figure 4.10B]**. Turning to residues interacting with the 3’ consensus flank, Arg220 and Lys221 demonstrated a loss of occupancy in the high-affinity structure suggesting an increase in sidechain dynamics in this flanking region **[Figure 4.12]**. No direct interactions with DNA bases occur in this region and increases in hydration with binding to high-affinity sequences may result from increased conformational dynamics suggesting an entropic component of affinity tuning [41]. Local B-factor increases observed in phosphate backbones of the high-affinity strand also suggest biased interactions between basic sidechains with phosphate groups in the low-affinity structure rather than dynamic contacts with the adjacent backbone or unobserved solvent mediated interactions.

4.4 Discussion

These crystallographic experiments gave rise to the highest-resolution structures of any DNA-bound ETS domain. Additionally, our determined structures represent some of the most highly resolved of transcription factors compared to those deposited in the RCSB Protein Data Bank to date. Structural similarity in the backbone dihedrals at the macroscopic level not only lends credence to what can visually be observed, but also provides concrete proof that DNA selectivity is born from the orientation and dynamics of the amino acid side chains. With the overall structure remaining unchanged across multiple experimental conditions, the differences in the amino acid conformations must play a role in DNA recognition and selectivity.

MD variances in Q226 occupancy gave insight into PU.1 divergent tunability relative to E226 orthologs. Variable occupancy between high- and low-affinity structures at this position as well as R220 and K221 demonstrated the effects of flanking sequence variability on sidechain dynamics which alter affinity. Additionally, R220 and K221 motility shifts between the two structures yielded further insight into the observable differences in hydration patterns when bound to high- and low-affinity sequences.

4.5 Methods

4.5.1 Nucleic Acids

Deoxynucleotides for co-crystallization were prepared by solid state synthesis and purified by reverse-phase HPLC by Integrated DNA Technologies (Coralville, IA). Lyophilized DNA was redissolved and annealed in Buffer H (10 mM HEPES and 0.15 M NaCl adjusted to pH 7.4 with NaOH). The molecular cloning of the wildtype ETS domain of human PU.1 (residues 165 to 270, termed ΔN165) has been previously described [25]. Point mutants of ΔN165 were generated following a PCR-based strategy and cloned into pET28b or pCDF-1b vectors (Novagen) as previously described [26].

4.5.2 Protein Purification

Heterologous over-expression was conducted in BL21(DE3)pLysS *Escherichia coli* and purification was performed as described previously [25]. In brief, cultures in LB medium were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at an OD_{600} of 0.6 for 16 hours at 23°C. Harvested cells were re-suspended in Buffer A (10 mM HEPES, pH 7.4, and 0.5 M NaCl) containing 1 mM PMSF and lysed by sonication. The lysate was cleared by centrifugation and loaded onto HiTrap SP column (Cytiva) equilibrated with Buffer A. After washing, the protein was eluted along a linear NaCl gradient in Buffer A under the control of a GE Akta Start instrument. Following characterization by SDS-PAGE and mass spectrometry [**Figures 4.5B and C**], the protein was further purified by gel-filtration on a HiLoad Superdex 75 (GE Healthcare) pre-equilibrated with buffer H (10 mM HEPES, pH 7.4, 0.15 M NaCl). Concentration was determined by UV absorption at 280 nm based on an extinction coefficient of 22,460 M^{-1} cm^{-1}.

4.5.3 Crystallization

Purified PU.1ΔN165 and duplex DNA was mixed at 400 µM each in Buffer H to yield a final concentration of 200 µM complex. The complex was subjected to two different screens (INDEX and JCSG) of 96 wells conditions each using the ART Robbins Gryphon Robot. The most promising conditions were then explored further by x-ray diffraction and a primary condition identified yielding high resolution data. Larger crystals were grown for 5 days by vapor diffusion at 298 K in a hanging drop comprised of a 1:1 mixture of protein: DNA complex with mother liquor containing 100 mM sodium acetate, pH 4.6, and 2.5% PEG 3350 [**Figure 4.5D**]. Prior to freezing, 2 µL of cryoprotectant solution containing 100 mM sodium acetate, 2.5% PEG 3350, and 20% glycerol was laid on top of the hanging drop and the well closed for 1 hour of incubation (4 µL total volume, 10% glycerol concentration). After 1 hour, crystals were transferred to the above 20% glycerol solution prior to flash freezing in liquid nitrogen.

4.5.4 X-ray Data Collection and Processing

X-ray diffraction data sets were obtained using a Dectris Eiger X16M detector SER-CAT at the Advanced Photon Source, USA on the ID-22 beamline at a fixed wavelength of 1.0000 Å with an oscillation angle of 0.25°. The diffraction data was processed using the XDS package [27] and was scaled using Aimless in the CCP4i2 package [28]. The complex co-crystallized in the $P2_1$ space group with a single complex per asymmetric unit, in accordance with a Matthews coefficient of $V_M = 2.33$ Å^3/Da [29]. Molecular replacement was performed using a previous PU.1 co-crystal complex (PDB: 1PUE_E) as the search coordinates in the PHENIX suite [30] via the maximum-likelihood procedures in PHASER. Refinement was then carried out using phenix.refine [30] and model building was performed with Coot [31]. Backbone geometry and secondary assignment for the protein was computed using

DSSP [32]. Sidechain dihedral angles were calculated using the bio3d library [33]. DNA helical parameters were computed using x3DNA [34]. Other measurements and selection procedures were performed in PyMOL.

4.5.5 Molecular Dynamics Simulations

Explicit-solvent simulations were performed with the Amber14SB/parmbsc1 forcefields [35] in the GROMACS 2020.2 environment. The refined co-crystal structure was used as initial coordinates of the wildtype PU.1/DNA complex as well as to template any point mutant. The solution NMR structure of unbound PU.1 was used for the free state. Each system was set up dodecahedral boxes at least 1.0 nm wider than the longest dimension of the solute, solvated with TIP3P water, and neutralized with Na^+ and Cl^- to 0.15 M. Electrostatic interactions were handled by particle-mesh Ewald summation with a 1 nm distance cutoff. All simulations were carried out at an *in silico* temperature and pressure of 298 K (modified Berendsen thermostat) [36] and 1 bar (Parrinello-Rahman ensemble). A timestep of 2 fs was used and H-bonds were constrained using LINCS. After the structures were energy-minimized by steepest descent, the *NVT* ensemble was equilibrated at 298 K for 1 ns to thermalize the system, followed by another 1 ns of equilibration of the *NPT* ensemble at 1 bar and 298 K. The final *NPT* ensemble was simulated without restraints for 0.5 μs, recording coordinates every 1 ps. Convergence of the trajectories were checked by RMSD from the energy-minimized structures, after adjustments for periodic boundary effects. Triplicate production runs were carried out using different random seeds in the velocity distribution. The final 100 ns of each equilibrated trajectory for analysis. For RMS fluctuation calculations, concatenated trajectories from the replicas were used. Other averages were expressed ± S.D.

4.6 Figures

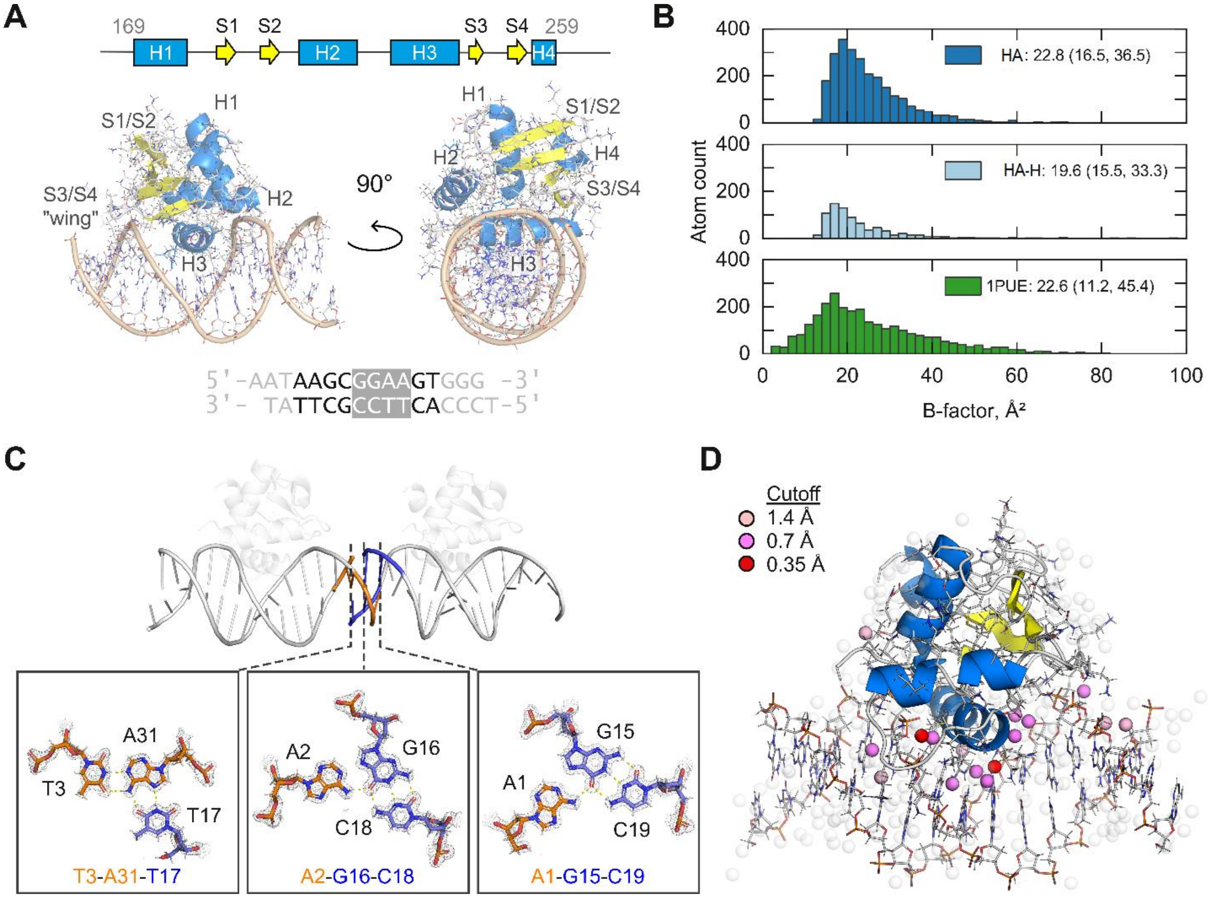


Figure 4.1 Comparative Analysis of a High-affinity DNA Complex of the ETS Domain of Human PU.1

A, Atomic coordinates of a single PU.1/high-affinity DNA complex. In the DNA sequence, positions contacted by protein within the complex are in bold. The core 5’-GGAA-3’ is inverted in color. ***B***, Distribution of crystallographic B-factors from the high-affinity structure (with and without emplaced hydrogens, HA-H) and 1PUE, which contains two non-equivalent complexes per asymmetric unit. ***C***, DNA-DNA crystal contacts in our structure. Mesh represents the 2Fo-Fc density at 1.5 σ. ***D***, Mapped water molecules (spheres) in the high-affinity structure, colored according to their spatial overlap with corresponding water in both non-equivalent complexes in 1PUE. Non-colored spheres represented water beyond a 1.4-Å cutoff.

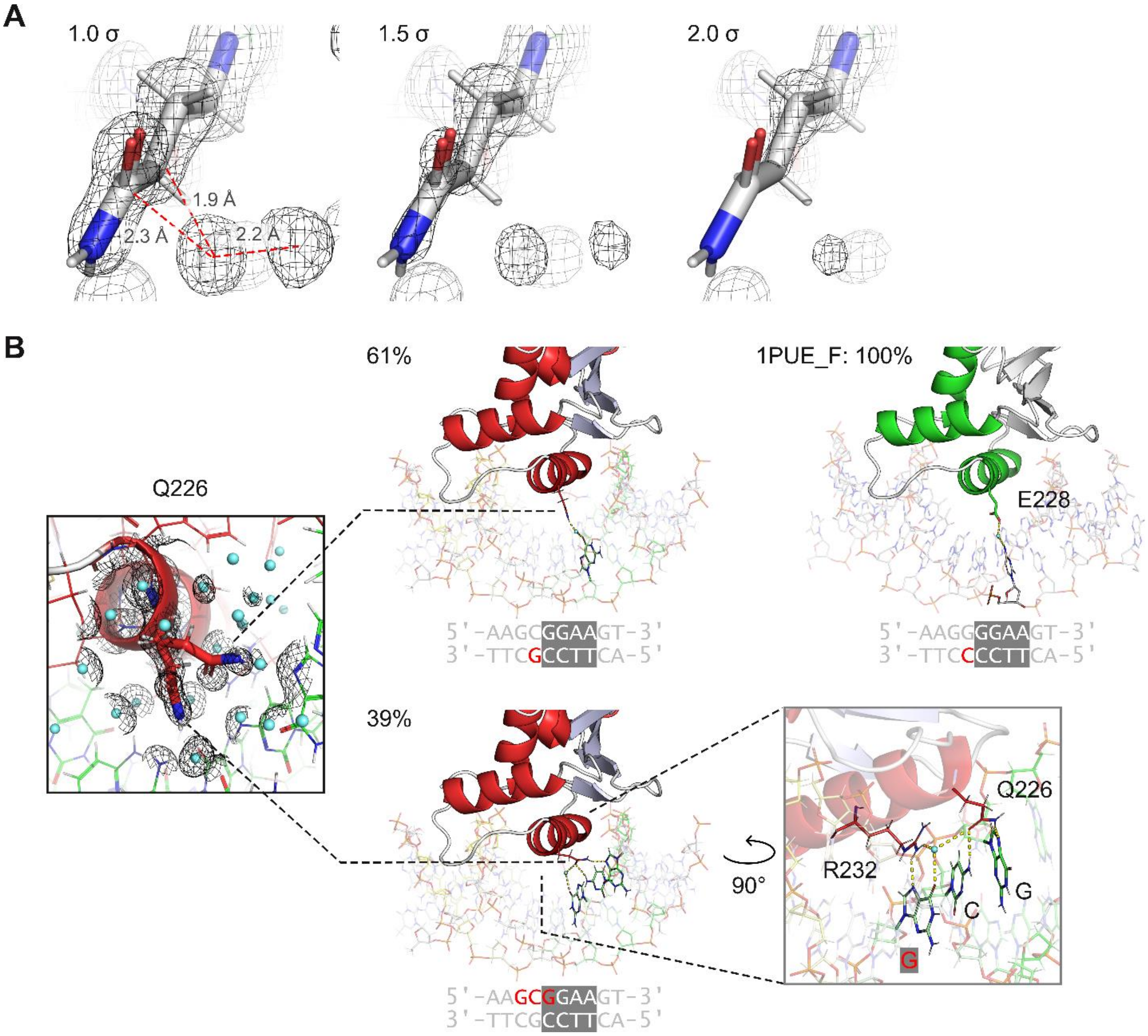


Figure 4.2 Alternate conformations of an evolutionarily divergent glutamine residue couples core and flanking sequence recognition in PU.1
A, The 2Fo-Fc map defining one conformation of Q226 at the indicated σ cutoffs. ***B***, Interactions of two fitted conformations of Q226 with the DNA target. The "Q226-up" conformation connects bases in both core (invert-colored text) and flanking positions (orange text) in a network involving R230, an absolutely conserved residue in the DBDs of ETS proteins. In contrast, the "Q226-down" position contacts only a proximal flanking base via a water-mediated contact, analogous to the glutamate-mutant co-crystallized in 1PUE. Positions contacted by Q226 are marked in black.

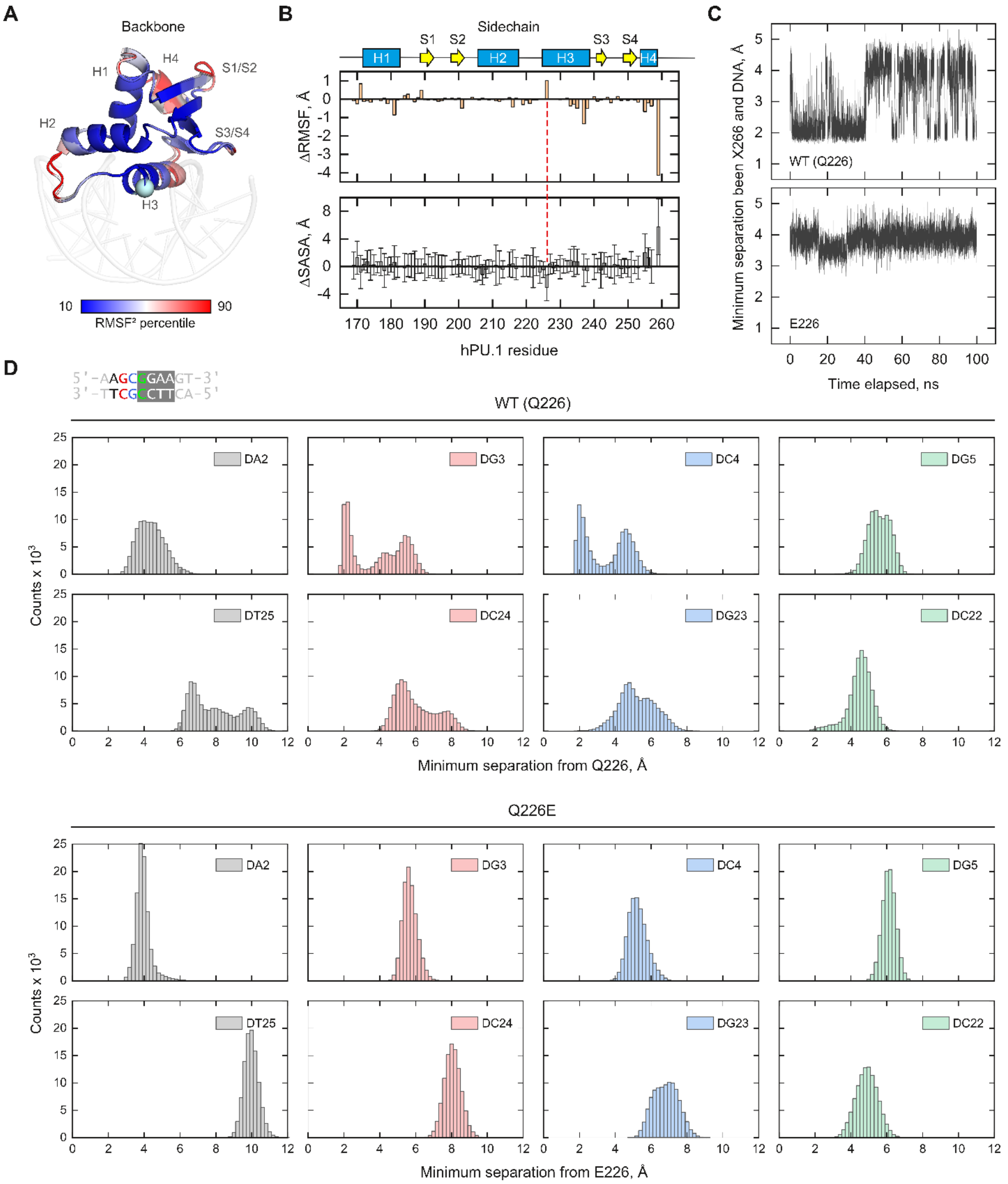


Figure 4.3 Dynamic consequences of an ancestral reversion in the high-affinity hPU.1/DNA complex

A, Backbone dynamics of the wildtype and Q226E simulational complexes. Each complex is colored from the 10th and 90th percentile in its respective spectrum of RMSF². A DNA is modeled to guide the orientation of the protein in the complex. *B*, Comparison of sidechain dynamics RMSF (heavy atoms only) and solvent-exposed surface area (SASA), expressed as the difference of wildtype minus Q226E. All RMSF values computed from the concatenated trajectories of triplicate simulations. ΔSASA are computed from time-averages ± SD. The red dashes indicate position 226. *C*, Minimum separation from the residues at position 226 to DNA over an illustrative 100-ns period. *D*, Minimum separation distances from the indicated DNA base positions for the trajectory shown in Panel C.

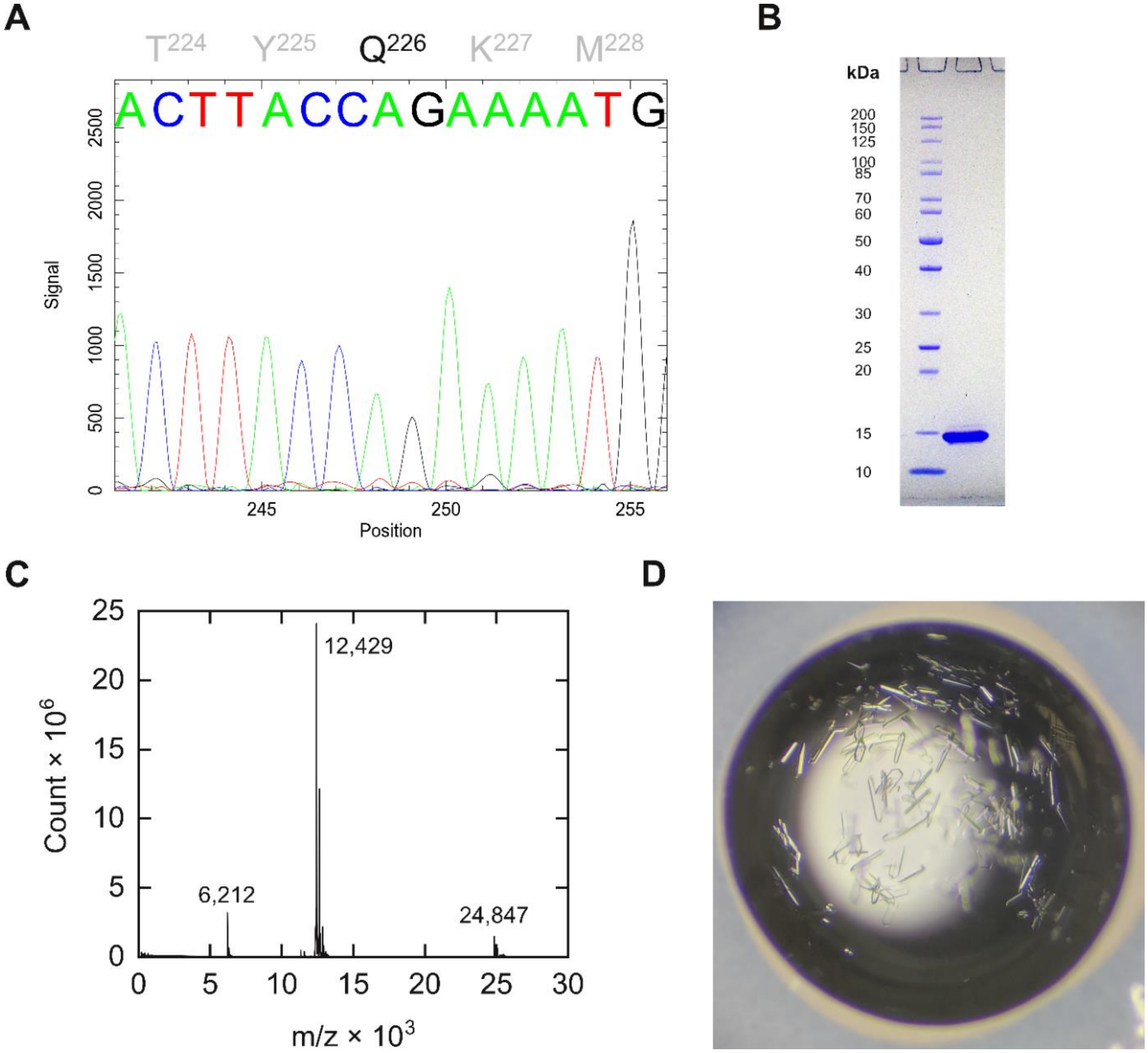


Figure 4.4 Qualification of the recombinant hPU.1 ETS domain used for co-crystallization with DNA

A, Sanger sequencing of the DNA construct encoding the hPU.1 ETS domain, focusing on the neighborhood of Q226. ***B***, SDS-PAGE analysis of 100 ng of purified protein. The gel was stained with Coomassie Blue. ***C***, MALDI-ToF mass spectrometric analysis. The expected mass is 12,406 Da. In addition to the primary peak, minor peaks corresponding to a $z = +2$ ion and a gas-phase dimer are also observed. ***D***, Micrograph of hPU.1 co-crystals with high-affinity DNA. The image was acquired by microscope of the 1:1 complex (400 µM) crystal in 100 mM sodium acetate pH 4.6 containing 2.5% PEG 3350 by the hanging drop vapor method.

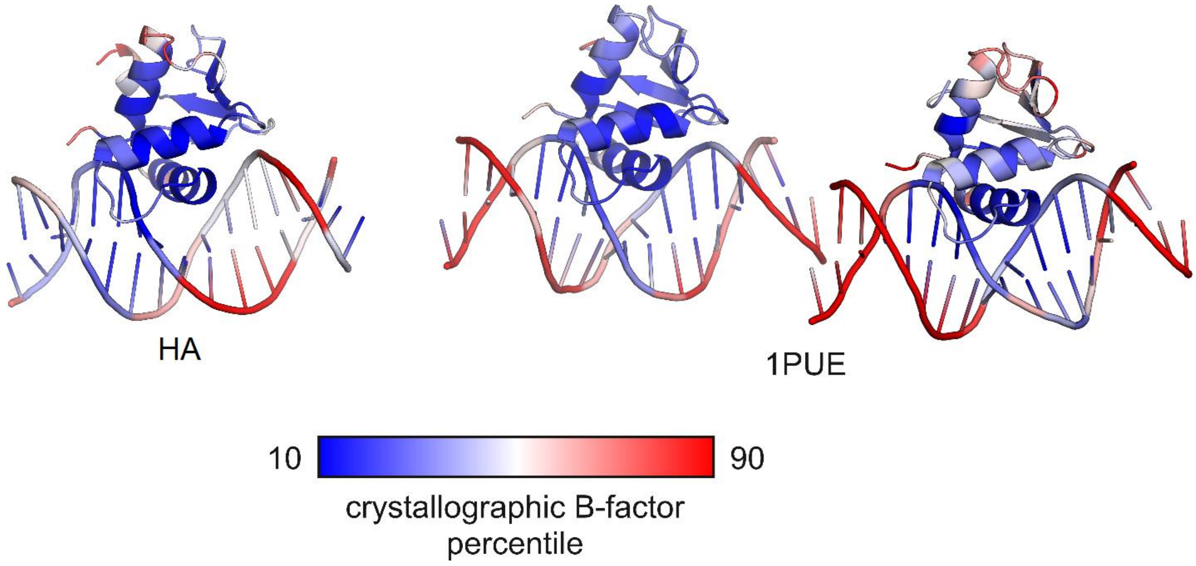


Figure 4.5 Crystallographic B-factor distribution in the high-affinity structure and 1PUE Cartoons of the high-affinity structure and the non-equivalent complexes in 1PUE are colored by their crystallographic B-factor on scale from their respective 10th to 90th percentile. The 10/90 inter-percentile range is 15.4 to 33.2 for HA, and 11.2 to 45.4 for 1PUE, respectively.

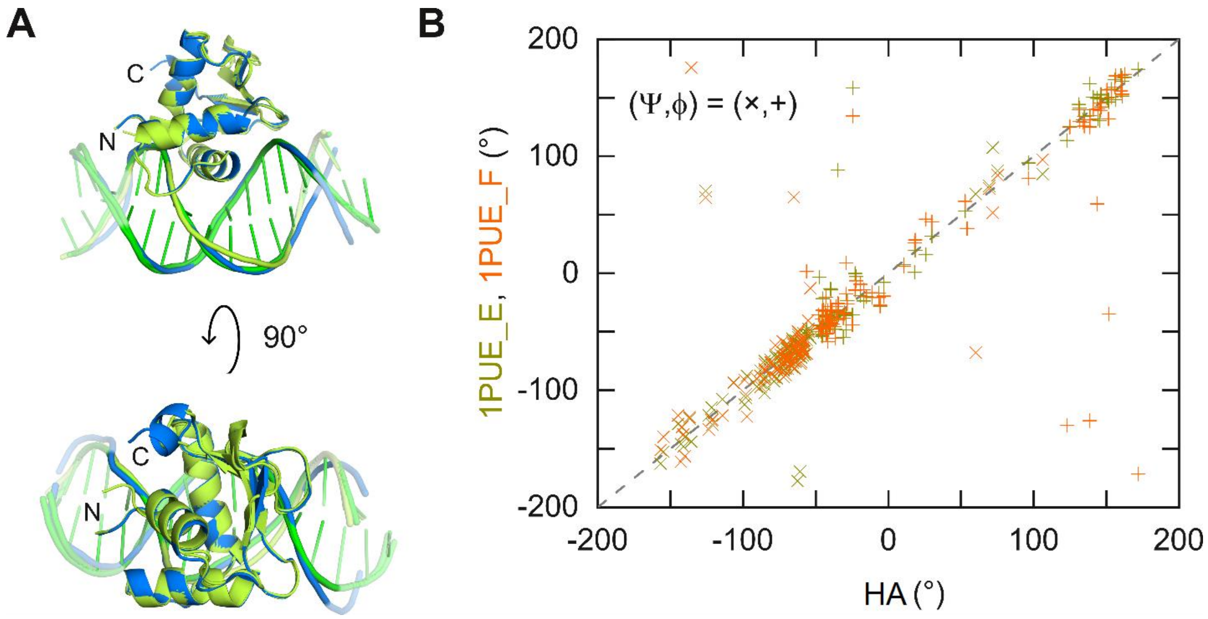


Figure 4.6 Comparison of backbone dihedral angles between the high-affinity structure and 1PUE
A, Superposition of the HA structure (blue) and 1PUE (green) complexes, aligned by the proteins. DNA positions no contacted by protein within the complex are rendered semi-transparent. *B*, ψ (×) and ϕ (+) were computed using DSSP. The asymmetric unit of 1PUE contained two non-equivalent complexes (chains E and F), both of which are used in the comparison. The dashed line is the identity function.

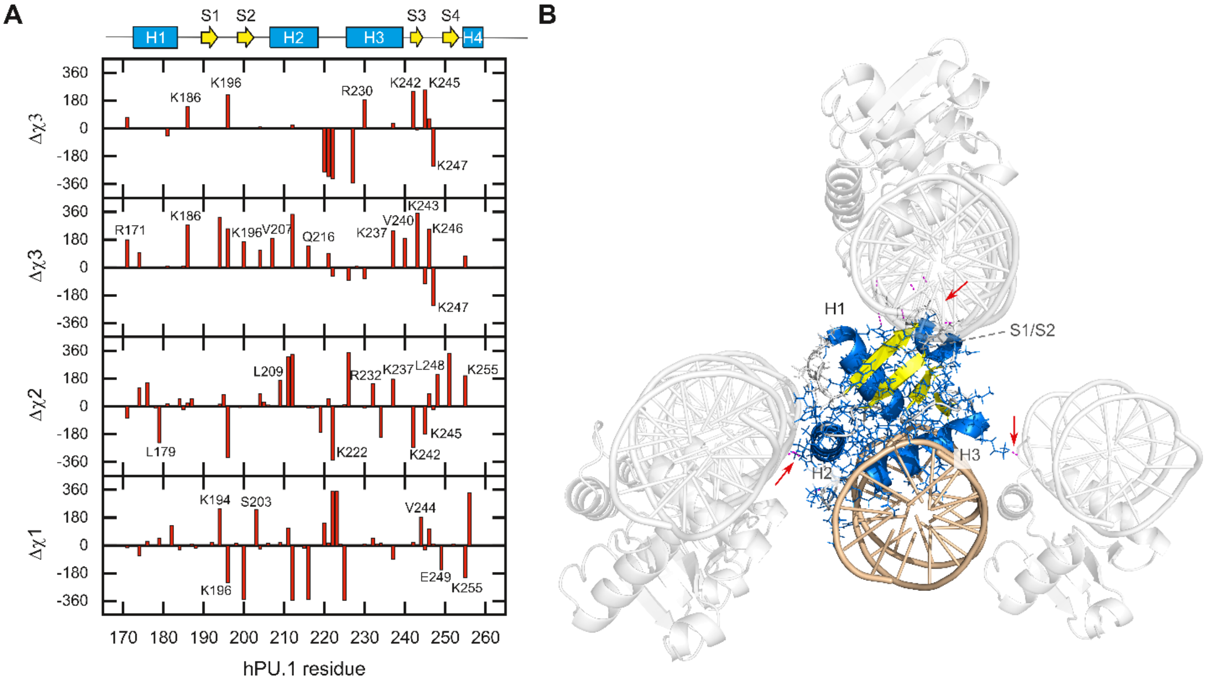


Figure 4.7 Comparison of sidechain dihedral angles between high-affinity bound and 1PUE structures

A, The mean values of corresponding residues in the two complexes in 1PUE were subtracted from their counterparts in the HA structure, and plotted against the residue numbers. Note that Δχ values near ±360° reflect similar rotameric states as 0° given their periodic nature. Accordingly, values around ±180° signify the greatest differences. *B*, Crystal contacts observed in HA. The secondary structure of one copy of the protein was colored according to the scheme in the cartoon in Panel A. Dashes (pointed by arrows) denote contacts between packed complexes within a cutoff of 4.5 Å.

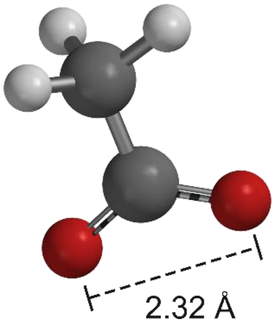


Figure 4.8 Optimized geometry of ground-state acetate in a low-dielectric environment Geometry optimization by DFT at the ωB97X-V/6-311++G** level of theory was performed using Spartan 18. An implicit solvent of dielectric ε = 8 was used to mimic typical electrostatic fields near duplex DNA [40].

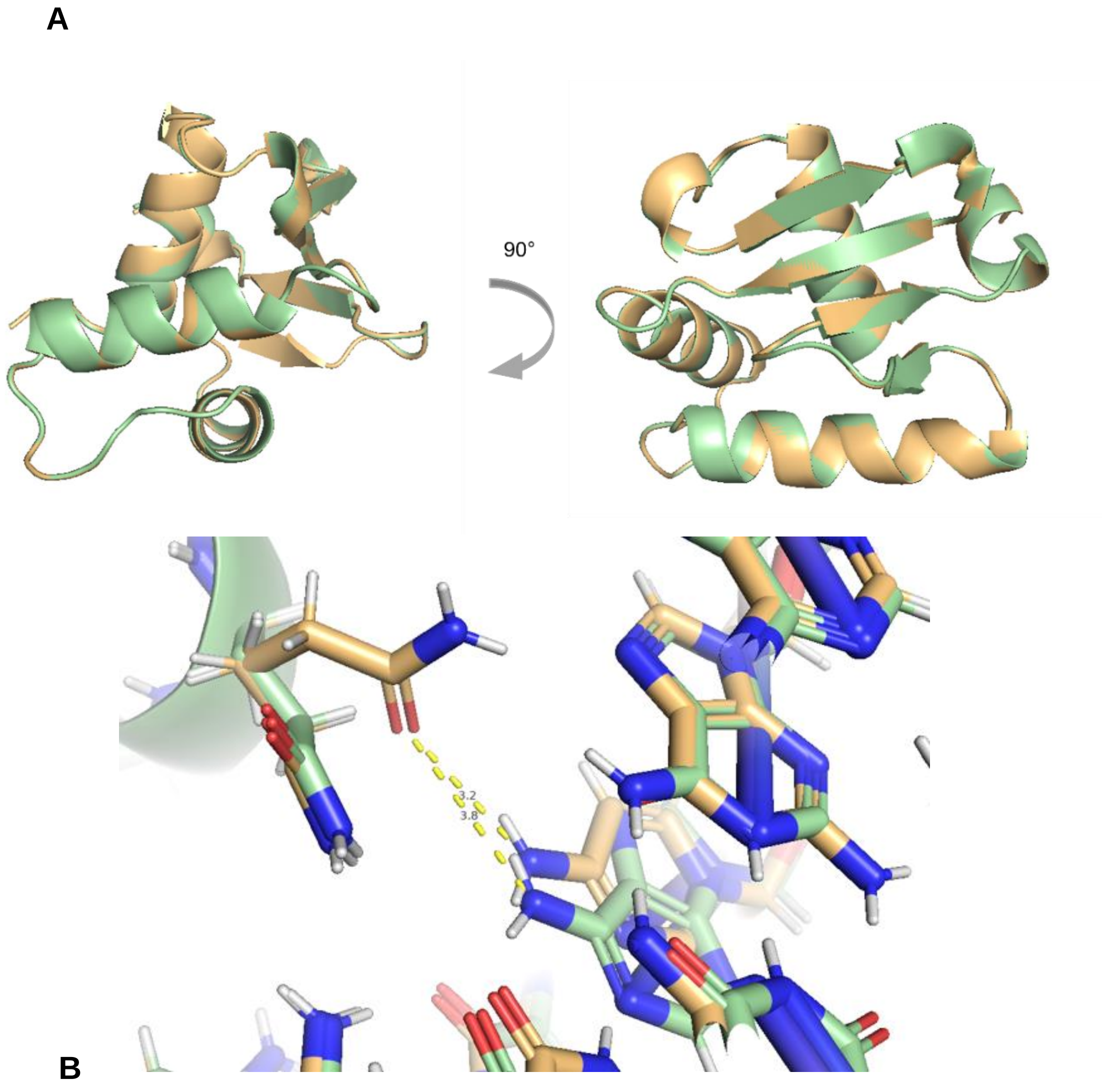


Figure 4.9 Protein alignment between high and low-affinity bound structures
A Alignment of the high-affinity structure (orange) with the low affinity structure (green). No conformational changes are observed in the protein backbone between these two structures suggesting the change in affinity results from sidechain conformational shift or dynamics in the DNA. **B** Alternative conformations and sidechain distance from dN7 amine groups at Q226

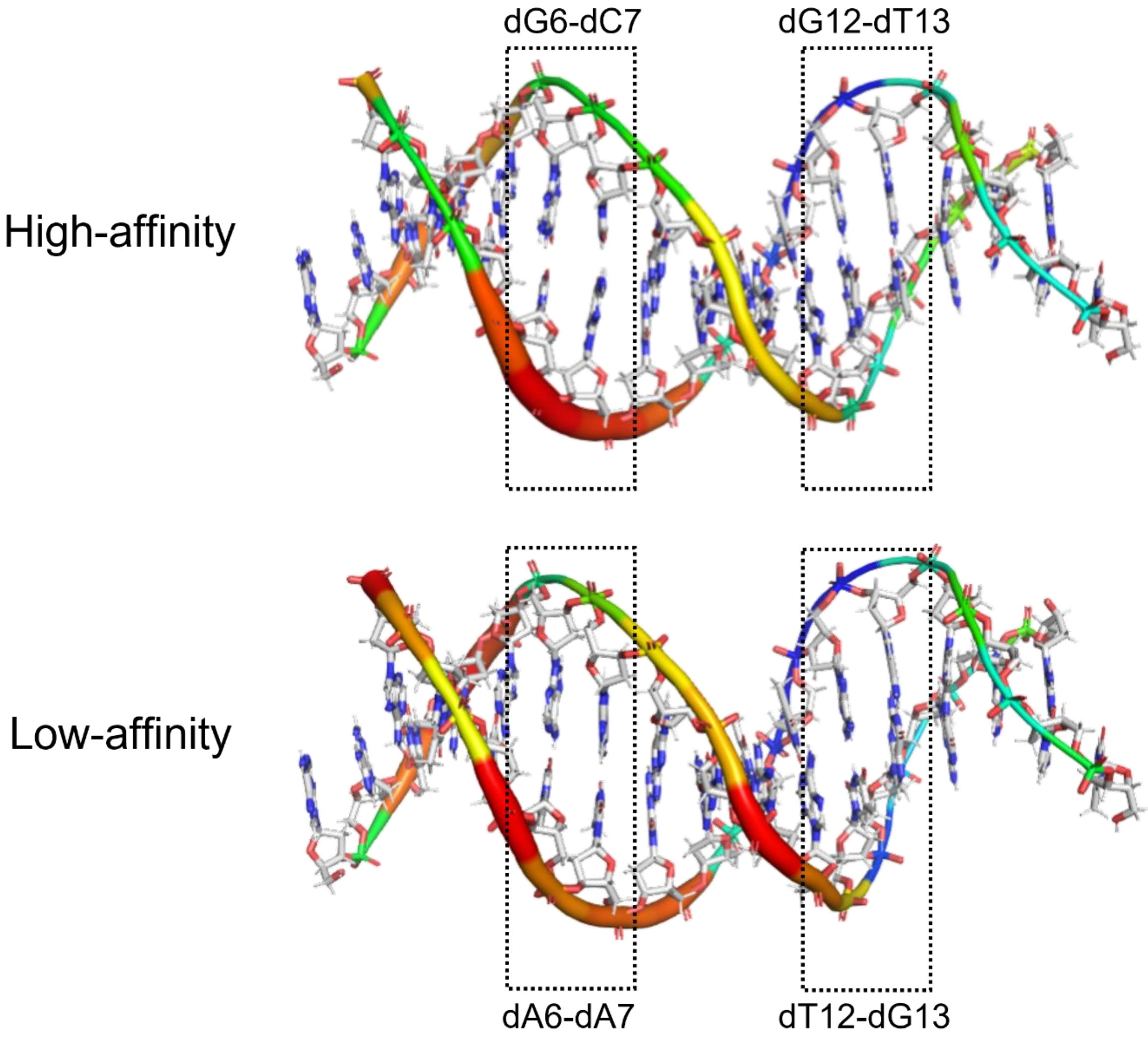


Figure 4.10 DNA B-factors of high and low-affinity DNA backbones
Variability in B-factor distribution is seen in the phosphate backbones of the consensus flanking regions. B-factor increases observed coincide with sidechain conformational variability promoting DNA contacts in these regions

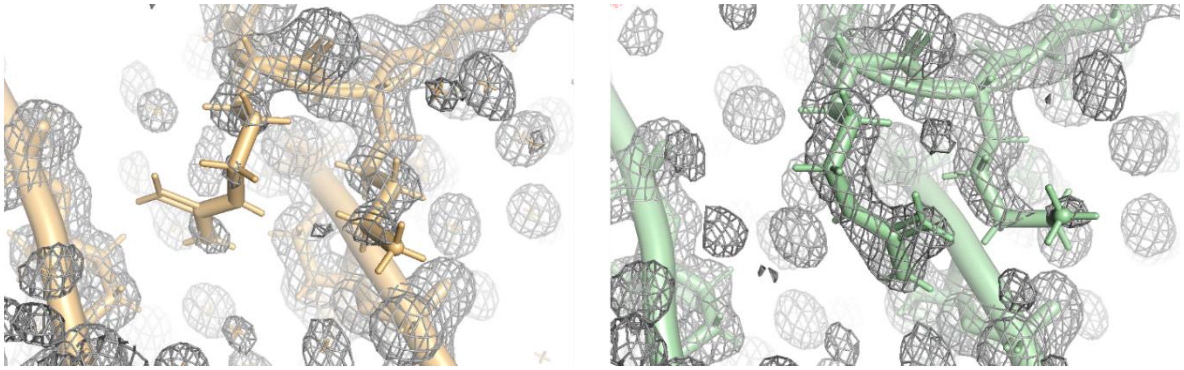


Figure 4.11 R220 and K221 sidechain conformations with 2F0-Fc map overlay at 1σ Increased occupancy of R220 in the low-affinity structure (right, green) suggests conformational conservation with hydrogen bonding to the phosphate backbone. Decreased occupancy is observed in the high-affinity structure (left, orange)

Table 4.1 Crystallographic Statistics for High-affinity and Low-Affinity PU.1 Complex Structures

	High-affinity bound complex (HA)	Low-affinity bound complex (LA)
5’ ---- 3’ DNA Sequence (overhang)	AATAAGCGGAAGTGGG	AATAAAAGGAAGTGGG
Wavelength (Å)	1.0000	1.0000
Resolution range (Å)	33.3 - 1.28 (1.326 - 1.28)	23.92 - 1.28 (1.326 - 1.28)
Space group	P 1 21 1	P 1 21 1
Unit cell (Å)	43.001 60.742 44.574	42.895 61.434 44.765
Unit cell (o)	90 116.714 90	90 117.231 90
Total reflections	195283 (19975)	101385 (10315)
Unique reflections	51827 (5190)	52198 (5243)
Multiplicity	3.8 (3.8)	1.9 (2.0)
Completeness (%)	98.11 (98.65)	98.00 (98.90)
Mean I/sigma(I)	11.84 (2.32)	16.98 (4.57)
Wilson B-factor (Å^2)	17.43	16.17
R-merge	0.06262 (0.4741)	0.04308 (0.2447)
R-meas	0.07385 (0.5506)	0.05051 (0.2858)
R-pim	0.03842 (0.2761)	0.02586 (0.1446)
CC1/2	0.992 (0.896)	0.996 (0.963)
CC*	0.998 (0.972)	0.999 (0.989)
Reflections used in refinement	51765 (5187)	52137 (5232)
Reflections used for R-free	1985 (211)	2504 (260)
R-work	0.1251 (0.1700)	0.1338 (0.1505
R-free	0.1510 (0.1985)	0.1550 (0.1725)
CC(work)	0.972 (0.959)	0.971 (0.970)
CC(free)	0.979 (0.948)	0.975 (0.946)
Number of non-hydrogen atoms	1724	1729
• macromolecules	1399	1385
• ligands	8	1
• solvent	320	343
Protein residues	91	91
RMS(bonds) (Å)	0.008	0.008
RMS(angles) (o)	1.09	1.15
Ramachandran favored (%)	97.75	97.75
Ramachandran allowed (%)	2.25	2.25
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Clashscore	3.53	1.19
Average B-factor (Å2)	25.81	24.77
• macromolecules	22.26	20.86
• ligands	49.43	32.48
• solvent	39.91	37.69

	Pair	HB#	WC?	δ, °		χ, °	
4	A-T	2	Y	111.7	120	-118.7	-109.8
5	A-T	2	Y	125.6	130.6	-115.1	-116.9
6	G-C	3	Y	104.1	120.2	-130.5	-129.6
7	C-G	3	Y	91.6	116.3	-147.3	-114.4
8	G-C	3	Y	147.1	105.5	-98.7	-126.5
9	G-C	3	Y	137.1	104.5	-110.9	-126.2
10	A-T	2	Y	110.6	103.5	-111.7	-120.9
11	A-T	2	Y	136	127	-109.4	-111.0
12	G-C	3	Y	135.7	119.6	-116.9	-163.1
13	T-A	2	Y	133.7	129.7	-107.8	-84.6
Average ± SD				121 ± 14		-119 ± 17	
1BNA				123 ± 21		-117 ± 14	

Table 4.2 Base pairing and torsion angle of PU.1-bound DNA positions
Parameters and interferences were provided by 3DNA. Comparison of the average sugar pucker (as indicated by the C5’-C4’-C3’-O3’ torsion angle δ) and glycosidic torsion angle χ is made with the classic Drew-Dickerson dodecamer, 5’-d(CGCGAATTCGCG)$_2$-3’ (PDB: 1BNA).

4.7 References

[1] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009;10:252-63.

[2] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172:650-65.

[3] Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science. 2007;316:1497-502.

[4] Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. Cell. 2013;152:327-39.

[5] Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS. Origins of specificity in protein-DNA recognition. Annu Rev Biochem. 2010;79:233-69.

[6] Graves BJ, Petersen JM. Specificity within the ets family of transcription factors. Adv Cancer Res. 1998;75:1-55.

[7] Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, et al. Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. EMBO J. 2010;29:2147-60.

[8] Hollenhorst PC, Jones DA, Graves BJ. Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic acids research. 2004;32:5693-702.

[9] Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem. 2011;80:437-71.

[10] Szymczyna BR, Arrowsmith CH. DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. The Journal of biological chemistry. 2000;275:28363-70.

[11] Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. The role of DNA shape in protein-DNA recognition. Nature. 2009;461:1248-53.

[12] Pio F, Ni CZ, Mitchell RS, Knight J, McKercher S, Klemsz M, et al. Co-crystallization of an ETS domain (PU.1) in complex with DNA. Engineering the length of both protein and oligonucleotide. The Journal of biological chemistry. 1995;270:24258-63.

[13] Kodandapani R, Pio F, Ni CZ, Piccialli G, Klemsz M, McKercher S, et al. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature. 1996;380:456-60.

[14] Hohaus S, Petrovick MS, Voso MT, Sun Z, Zhang DE, Tenen DG. PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene. Molecular and cellular biology. 1995;15:5830-45.

[15] Nagy M, Chapuis B, Matthes T. Expression of transcription factors Pu.1, Spi-B, Blimp-1, BSAP and oct-2 in normal human plasma cells and in multiple myeloma cells. British journal of haematology. 2002;116:429-35.

[16] Rao S, Garrett-Sinha LA, Yoon J, Simon MC. The Ets factors PU.1 and Spi-B regulate the transcription in vivo of P2Y10, a lymphoid restricted heptahelical receptor. The Journal of biological chemistry. 1999;274:34245-52.

[17] Satoh J-i, Asahina N, Kitano S, Kino Y. A Comprehensive Profile of ChIP-Seq-Based PU.1/Spi1 Target Genes in Microglia. Gene Regulation and Systems Biology. 2014;8:127-39.

[18] Garrett-Sinha LA, Su GH, Rao S, Kabak S, Hao Z, Clark MR, et al. PU.1 and Spi-B are required for normal B cell receptor-mediated signal transduction. Immunity. 1999;10:399-408.

[19] Moreau-Gachelin F, Tavitian A, Tambourin P. Spi-1 is a putative oncogene in virally induced murine erythroleukaemias. Nature. 1988;331:277-80.

[20] Ridinger-Saison M, Boeva V, Rimmele P, Kulakovskiy I, Gallais I, Levavasseur B, et al. Spi-1/PU.1 activates transcription through clustered DNA occupancy in erythroleukemia. Nucleic acids research. 2012;40.

[21] Roos-Weil D, Decaudin C, Armand M, Della-Valle V, Diop MK, Ghamlouch H, et al. A Recurrent Activating Missense Mutation in Waldenstrom Macroglobulinemia Affects the DNA Binding of the ETS Transcription Factor SPI1 and Enhances Proliferation. Cancer Discov. 2019;9:796-811.

[22] Seki M, Kimura S, Isobe T, Yoshida K, Ueno H, Nakajima-Takagi Y, et al. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. Nat Genet. 2017;49:1274-81.

[23] Degnan BM, Degnan SM, Naganuma T, Morse DE. The ets multigene family is conserved throughout the Metazoa. Nucleic acids research. 1993;21:3479-84.

[24] Poon GM, Macgregor RB, Jr. Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. J Mol Biol. 2003;328:805-19.

[25] Xhani S, Lee S, Kim HM, Wang S, Esaki S, Ha VLT, et al. Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. Sci Adv. 2020;6:eaay3178.

[26] Albrecht AV, Kim HM, Poon GMK. Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. Nucleic acids research. 2018;46:10577-88.

[27] Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010;66:125-32.

[28] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67:235-42.

[29] Kantardjieff KA, Rupp B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 2003;12:1865-71.

[30] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66:213-21.

[31] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66:486-501.

[32] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577-637.

[33] Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22:2695-6.

[34] Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic acids research. 2003;31:5108-21.

[35] Ivani I, Dans PD, Noy A, Perez A, Faustino I, Hospital A, et al. Parmbsc1: a refined force field for DNA simulations. Nat Methods. 2016;13:55-8.

[36] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126:014101.

[37] Poon GM. Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. The Journal of biological chemistry. 2012;287:18297-307.

[38] Wang S, Linde MH, Munde M, Carvalho VD, Wilson WD, Poon GM. Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the

ETS family transcription factors Ets-1 and PU.1. The Journal of biological chemistry. 2014;289:21605-16.

[39] Xhani S, Esaki S, Huang K, Erlitzki N, Poon GM. Distinct Roles for Interfacial Hydration in Site-Specific DNA Recognition by ETS-Family Transcription Factors. J Phys Chem B. 2017;121:2748-58.

[40] Cuervo A, Dans PD, Carrascosa JL, Orozco M, Gomila G, Fumagalli L. Direct measurement of the dielectric polarization properties of DNA. Proc Natl Acad Sci U S A. 2014;111:E3624-30.

[41] Morgunova1 E., Yin Y.,Das P.K., Jolma A., Zhu F., Popov A., Xu Y., Nilsson L.,Taipale J. Two distinct DNA sequences recognized by transcription factors represent enthalpy and entropy optima. *eLife* 2018;7:e32963

5 SUMMARY AND CLOSING REMARKS

5.1 Summary

Utilization of molecular probes to perturb protein crystal structures has been demonstrated as a robust method for obtaining information about protein structure-function relationships. In sections 2 through 4 eight novel crystal structures, six of which included molecular probes, were presented. These structures yielded new insights into structure-function relationships in their respective systems. Additionally, a new method for structural perturbation was demonstrated by rapid freezing of crystals in liquid oxygen.

In section 2, incorporation of molecular oxygen into hemoglobin crystals by rapid freezing in LO_2 was examined as a potential method for observing channels which comprise the routes for molecular oxygen to reach the protein heme centers. Observation of crystallographic B-factor increases (relative to an untreated control structure) at previously postulated diffusion channels was observed, suggesting a rapid influx of molecular oxygen through these regions. This demonstrated that even during freezing, large quantities of molecular oxygen were able to permeate the crystal lattice. Further substantiation of the observed changes were indicated following cross-referencing with a database cataloging hemoglobin mutations. This cross-reference showed that amino acid variance in these regions can often result in the clinical presentation of disease states such as anemia and thalassemia from which a disruption of diffusion pathways can be inferred. Ultimately the experiment presented the successful development of an accessible and novel method for crystallographers to examine proteins which utilize molecular oxygen.

In section 3, the MDM2 RING domain was demonstrated to be a high value target for potential therapeutic development for the treatment of various cancers due to its role as an E3 ligase mediating the ubiqutination of p53. Several compounds which bind the MDM2 RING domain have demonstrated inhibitory activity against E3 ligase activity towards p53, however, their exact binding sites and mechanisms of action had not been established. Crystallographic examination of the MDM2 RING domain following the incorporation of inulanolide A and MA242 via extended soaking demonstrated that these two small molecule inhibitors share a common binding site adjacent to the C-terminal residues Tyr489 and Phe490. Examination of the roles of these residues by other groups has demonstrated that knockout mutants possess little to no E3 activity suggesting that this site may be promising for the disruption of the anti-p53 activity of MDM2. Comparison of determined structures with those of UbcH5B-Ub-MDM2 RING ternary complexes demonstrated that steric disruption of Ubiquitin-MDM2 RING interfacing is likely and decreases in the autoubiquitination rates of MDM2 RING domain in the presence of both of these compounds served as a biochemical validation of interfacial disruption.

Section 4 yielded insights into the role of flanking sequence effects on the affinity and resulting transcriptional activation levels of the ETS family transcription factor PU.1. Co-crystallization of the PU.1 DNA binding domain with high- and low-affinity DNA sequences in a novel condition resulted in high resolution crystal structures which demonstrated subtle variances in conformation at residues Q226 and R220 as well as B-factor changes in the DNA phosphate backbone between the two structures. MD simulations examining sidechain dynamics at the Q226 position also suggested a potential mechanism through which PU.1 achieves divergent expression relative to other ETS family transcription factors through dynamic interaction with the N7 position of a 5’ flanking purine residue.

5.2 Future Directions

While the general story regarding crystallographic probing of hemoglobin using LO_2 is complete, the methodology can still be utilized in novel ways for the purposes of examining gas channels or for trapping reactive intermediates via *in crystallo* freeze quenching. Hemoglobin was an excellent model system, but there are many other examples of proteins and enzymes which utilize O_2 which could present interesting structures following LO_2 treatment. Highly reactive and short-lived species such as Fe (IV) and peroxyflavin catalytic intermediates may be promising targets as they have both been observed by stopped-flow spectrophotometry methodologies [1,2]. If grown in anaerobic conditions, crystals containing the necessary iron/flavin cofactors may be among the most promising targets for future studies attempting to examine the structure of these short-lived species. Additionally, neutron diffraction experiments utilizing the method described may be able to differentiate waters from O_2 molecules in LO_2 soaked structures and allow for the determination of specific gas binding regions in treated protein crystals [3].

The structural studies on the inhibitors inulanolide A and MA242 yielded two potential future directions for research. The first is simply a structure-based drug design approach to the development of novel InuA and MA242 analogues which inhibit MDM2. With the establishment of both the conditions for protein crystallization as well as a method of compound incorporation, many other small molecules could be examined structurally, biochemically, and *in vitro* for the advancement of cancer treatment potential against MDM2 specifically. However, if dual inhibition of NFAT1 and MDM2 is the goal, additional work must still be completed to examine how the binding of the small molecules to NFAT1 results in the inhibition of the *mdm2* oncogene. With the combined information from both structures, potential optimization towards both targets could be kept in mind during subsequent analog synthesis [4].

Lastly, the determined PU.1 structures gave much insight into past observations of flanking sequence effects on both binding affinity and transcriptional regulation, but there is much more information that can be obtained crystallographically. Having established a robust crystallization condition, additional sequences of DNA could be utilized to further examine regulatory mechanisms of PU.1 transcriptional activity such as DNA-methylation [5]. Alternate sequences may also be utilized to observe non-specific binding and gain insight into how PU.1 scans for target sequences and self regulates through dimerization [6].

5.3 References

[1] Lee, S., Shin, S., Li, X., & Davidson, V. L. (2009). Kinetic mechanism for the initial steps in MauG-dependent tryptophan tryptophylquinone biosynthesis. *Biochemistry*, *48*(11), 2442–2447.

[2] Gonzalez-Osorio, L., Luong, K., Jirde, S., Palfey, B. A., & Vey, J. L. (2016). Initial investigations of C4a-(hydro)peroxyflavin intermediate formation by dibenzothiophene monooxygenase. *Biochemical and biophysical research communications*, *481*(1-2), 189–194.

[3] Phillips, S., Schoenborn, B. Neutron diffraction reveals oxygen–histidine hydrogen bond in oxymyoglobin. *Nature* **292,** 81–82 (1981).

[4] Qin JJ, Wang W, Voruganti S, Wang H, Zhang WD, Zhang R. Identification of a new class of natural product MDM2 inhibitor: In vitro and in vivo anti-breast cancer activities and target validation. Oncotarget. 2015;6(5):2623-40.

[5] Stephens, D. C., & Poon, G. M. (2016). Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. *Nucleic acids research*, *44*(18), 8671–8681.

[6] Xhani, S., Lee S., Kim H-M., Wang, S., Esaki, S., Ha V.L.T, Khanezarrin M., Fernandez G.L., Albrecht, A.V., Aramini J.M., Germann, M.W., Poon, M.K.W. (2020) Intrinsic disorder controls two functionally distinct dimers of the master transcription factor PU.1. *Science Advances*, *6*(8)