

DISSERTATION

THE BIOPHYSICAL, BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF
POLY(ADP-RIBOSE) POLYMERASE-1 (PARP-1) AND ITS COMPLEXES WITH DNA-DAMAGE
MODELS AND CHROMATIN SUBSTRATES

Submitted by

Nicholas James Clark

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2013

Doctoral Committee:

Advisor: Karolin Luger

Susan Bailey

Jennifer DeLuca

Jeffrey C. Hansen

Robert Woody

Copyrighted by Nicholas J. Clark 2013

All Rights Reserved

ABSTRACT

THE BIOPHYSICAL, BIOCHEMICAL AND STRUCTURAL CHARACTERIZATION OF POLY(ADP-RIBOSE) POLYMERASE-1 (PARP-1) AND ITS COMPLEXES WITH DNA-DAMAGE MODELS AND CHROMATIN SUBSTRATES

Eukaryotic DNA is highly dynamic and must be compacted and organized with the help of cellular machines, proteins, into 'heterochromatin' state. At its basic level, chromatin is comprised of spool-like structures of protein complexes termed histones, which bind and organize DNA into larger fibrous structures. Cellular processes like transcription and DNA-damage repair require that chromatin be at least partially stripped of its protein components, which in turn allows for complete accessibility by DNA-repair or transcription machinery. A number of protein factors contribute to chromatin structure regulation. Poly(ADP-ribose) Polymerase-1 (PARP-1) is one of these proteins that exists in all eukaryotic organisms except for yeast. In its inactive form, it compacts chromatin, but performs its chromatin-opening function by covalently modifying itself and other nuclear proteins with long polymers of ADP-ribose in response to DNA damage. Thus, it also serves as a first responder to many types of DNA damage. The highly anionic polymers serve to disrupt protein-DNA interactions and thus allow for the creation of a temporary euchromatin structure.

Contained herein are investigations aimed at addressing key questions regarding key differences between the interactions of PARP-1 and chromatin and its DNA-damage substrates. Our experiments show that human PARP-1 interacts with and is enzymatically activated to a similar level by a variety of different DNA substrates. In terms of chromatin, it appears that PARP-1 fails to interact with nucleosomes that do not have linker DNA. PARP-1 most effectively interacts with chromatin by simultaneously binding two DNA strands through contacts made by its two N-terminal Zn-finger domains. Small-Angle X-ray (SAXS) and Neutron Scattering (SANS) and molecular dynamics (MD)

experiments were combined with biophysical and biochemical studies to better describe the structural effects of DNA binding on PARP-1. The average solution structure of PARP-1 indicates that the enzyme is a monomeric, non-spherical, elongated molecule with a radius of gyration (R_g) of $\sim 55\text{\AA}$. The DNA-bound form of PARP-1 is also monomeric and binding DNA causes the molecule to become more elongated with an average R_g of $\sim 80\text{\AA}$.

DEDICATION

I dedicate this dissertation to my mother who never got to see this document but was the primary inspiration behind its existence. I could not have completed my PhD without my family who was always so supportive. Thank you Aaron and Dad for your emotional and physical support especially during the cross-country moves that you made possible and memorable. To Amanda, Andrea and Angela for the laughs that kept me sane. To my grandparents who encouraged me to take the high road and when that the high road brought me near Wyoming, always had a place for me to stay. To my Aunts Meg and Ann for being positive examples with graduate degrees. To my many mentors, colleagues and scientist friends throughout my undergraduate and graduate careers, thank you for taking the time to help me become a better scientist. To my many friends over the years from which I learned a lot of life lessons, hobbies, humility and to occasionally take breaks and enjoy life.

To Sayward my adoring and supportive wife, without you I could not have succeeded at my PhD let alone my life. Your warmth and understanding cannot be matched. Thanks for the surprise birthday adventures during graduate school, the Colorado fourteeners and the national park trips. Also, a special thank you to both you and Wendy for the extra push to finish this document.

TABLE OF CONTENTS

| | |
|---|----|
| Abstract..... | ii |
| Dedication..... | iv |
| Table of Contents..... | v |
| Chapter 1: Review of Literature | |
| 1.1 Poly-ADP-ribose..... | 1 |
| 1.2 The PARP-superfamily..... | 5 |
| 1.3 PARP-1, A Structural and Functional Description..... | 7 |
| 1.4 PARP-1 in DNA Damage Repair..... | 15 |
| 1.4 PARP-1 in Chromatin and Transcription..... | 20 |
| 1.5 The Role of PARP-1 in Chromatin Architecture..... | 24 |
| 1.6 Specific Aims..... | 28 |
| Chapter 2: Alternative Binding Modes of Poly(ADP-ribose) Polymerase-1 to Free DNA and Nucleosomes | |
| 2.1 Chapter Overview..... | 31 |
| 2.2 Summary..... | 31 |
| 2.3 Introduction..... | 32 |
| 2.4 Experimental Procedures..... | 35 |
| 2.5 Results..... | 39 |
| 2.6 Discussion..... | 49 |
| 2.7 Acknowledgements..... | 54 |
| Chapter 3: Inferential Structure Determination of Multi-Domain Proteins from Small-Angle X-ray Scattering Data. | |
| 3.1 Chapter Overview..... | 55 |
| 3.2 Summary..... | 56 |

| | |
|---|-----|
| 3.3 Introduction..... | 57 |
| 3.3 Methods and Materials..... | 58 |
| 3.4 Results..... | 66 |
| 3.5 Discussion..... | 81 |
| 3.6 Acknowledgements..... | 85 |
| 3.7 Supporting Information..... | 86 |
| Chapter 4: Structural and Biophysical Studies of the Human PARP-1 in Complex with Damaged DNA | |
| Lilyestrom W., van der Woerd, M., <u>Clark, N.</u> , Luger, K., (2009) <i>J. Mol. Biol.</i> 395(5) 983-994 | |
| 4.1 Chapter Overview..... | 94 |
| 4.2 Summary..... | 94 |
| 4.3 Introduction..... | 95 |
| 4.4 Methods and Materials..... | 97 |
| 4.5 Results..... | 101 |
| 4.6 Discussion..... | 115 |
| 4.7 Acknowledgements..... | 118 |
| 4.8 Supporting Information..... | 119 |
| Chapter 5: The Structural Basis for the Allosteric Activation of PARP-1 by Small Oligonucleotide DNA Substrates | |
| 5.1 Chapter Overview..... | 125 |
| 5.2 Summary..... | 125 |
| 5.3 Introduction..... | 126 |
| 5.4 Methods and Materials..... | 130 |
| 5.5 Results..... | 139 |
| 5.6 Discussion..... | 177 |
| 5.7 Supporting Materials..... | 184 |

Chapter 6: Summary of Results

6.1 Discussion.....188

Chapter 7: Future Directions

7.1 Chapter Summary.....190

7.2 Significance.....191

7.3 Preliminary Data.....191

7.4 Experimental Plan.....194

References.....197

Appendix: X-ray crystallographic work as a contributing author.....212

Watanabe, S., Resch, M., Lilyestrom, W., Clark, N., Hansen, J. C., Peterson, C. & Luger, K. Structural characterization of H3K56Q nucleosomes and nucleosomal arrays. *Biochim. Biophys. Acta.* 1799, 480–486.

A.1 Appendix Overview.....212

A.2 Summary.....213

A.3 Introduction.....213

A.4 Methods and Materials.....215

A.5 Results.....217

A.6 Discussion.....224

A.7 Acknowledgements.....227

Chapter 1

Introduction

1.1 Poly-ADP-ribose

Many of the molecules that are essential for life exist as polymers. Nucleic acids and proteins are examples of important biological polymers. It was not until the mid-twentieth century that the real structural understanding of these complex biological chemistries began to take shape at the atomic level anyway. Advances in X-ray sources and X-ray diffraction available to researchers in the early 1950s, allowed Franklin and Gosling to obtain the fiber diffraction cross-section of DNA [1]. The resulting diffraction pattern led to the description of the molecular structure of B-form DNA by Watson and Crick [2]. A short time after the structure of DNA was deduced, the structure of myoglobin, the first protein structure, was published [3]. A novel polymer consisting of repeating units of ADP-ribose was first described just a few years later during the early 1960s. It was not known at the time that the discovery of ADP-ribose polymers would prove to have importance in many critical cellular pathways.

Extracts of chicken liver cell nuclei contained formations of what was initially thought to be polyadenylation (poly-A). In that study it was shown that a nuclear enzyme was responsible for the synthesis of a large quantity of the “poly-A”. It turned out that the synthetic reaction depended on both the presence of DNA and nicotinamide mononucleotide (NMN), rather than the presumptive substrate ATP [4]. The discovery of chains of poly-A that were synthesized in the absence of nucleotide triphosphates suggested a novel synthetic pathway. Additionally, the polymer they described was unlike other nucleic acids in that it had resistance to RNase and DNase activities. This suggested the existence of a third type of nucleic acid. Further support for the existence of a new type of nucleic acid came from the fact that ADP-ribose polymers were often purified with DNA and RNA. Later, it was reported that depending on the purification scheme used to purify the poly-A, the polymers of heterogeneous sizes would or would

not co-purify with nucleic acids, as reviewed by H. Hitz and P. Stone [5]. Shortly after the discovery of poly-A, several groups described the structure of the poly-A as a novel type of polymer consisting of repeating units of adenosine diphosphate-ribose, ADP-ribose [6-8]. The monomeric units are comprised of two ribose molecules and two phosphate moieties and the units can be thought of chemically as NAD⁺ molecules without a nicotinamide group (**Figure 1.1**).

Chambon *et al.* were correct in describing the novel reaction requirements, and later the fundamental structural elements of the polymer, but were incorrect in naming the NMN molecule as the starting molecule or building block of the polymer. The energy rich substrate for the previously unknown-enzyme catalyzed synthesis of chains of ADP-ribose is nicotinamide adenine dinucleotide (NAD⁺). A careful analysis, using doubly labeled ATP, NMN and NAD⁺ compounds allowed for the elucidation of the proper starting material [7, 8]. In the communication by Nishizuka *et al.* it was also shown that DNA had to be present in the reaction for polymer formation to occur [7]. Clearly, Chambon *et al.*, were on the right track but their experimental design may have overlooked the presence of other enzymes specifically, NAD⁺ pyrophosphorylase. NAD⁺ pyrophosphorylase was no doubt present in the nuclear enzyme preparations used in their experiments. NAD⁺ pyrophosphorylase is known to combine radioactive NMN with ATP to form the radioactive NAD⁺. Combining those results with insight from their own experiments, Nishizuka *et al.* were able to confirm the need for NAD⁺ to be present for the enzymatic formation of the polymers [7].

The discovery and early characterization of the poly-ADP-ribose lead some to speculate about the cellular purpose for such polymers. Upon describing what was thought to be poly-A, Chambon *et al.* hypothesized that the polymer served as a form of storage or containment for energy rich molecules [4]. Alternatively, NAD-derived chains of poly-ADP-ribose were proposed by Nishizuka *et al.* to be a formation of energetically costly NAD⁺ molecules bound in an irreversible manner and contained within cellular nuclei. Intriguingly, the authors calculated the rate of NAD⁺ consumption, during the synthesis of poly-ADP-ribose, and that rate seemed to exceed the rates of NAD⁺ formation by the two known pathways described by Chambon *et al.* and represented an unsustainable rate of NAD⁺ consumption [6, 7].

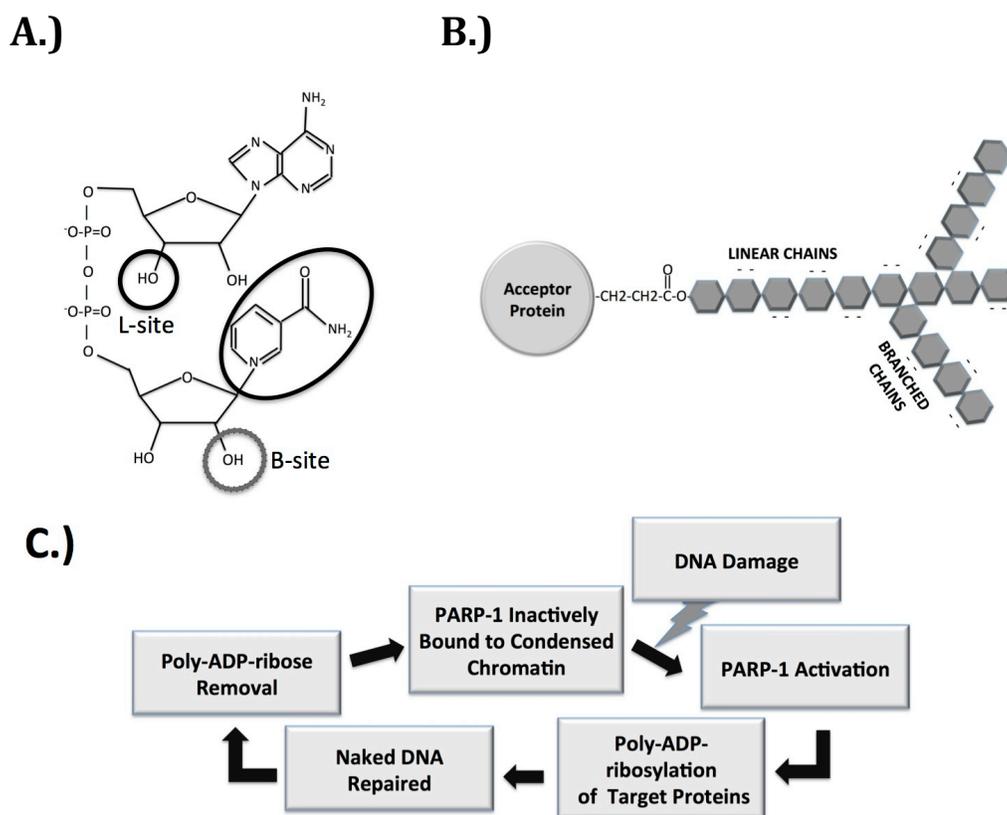


Figure 1.1. Nicotinamide Adenine Dinucleotide (NAD⁺) and poly-ADP-ribose (PAR). A.) the chemical structure of NAD⁺. The orange oval highlights the nicotinamide portion of the molecule, this group is a leaving group during the ADP-ribosylation reaction. The black circle represents the acceptor group for the successive addition of more ADP-ribose moieties during the synthesis of linear PAR chains. L-site and B-sites represent favorable acceptor sites for linear and branched chain formation, respectively. B.) PAR formation depends on the initial ADP-ribosyl modification of surface exposed amino acids on target proteins, in this case the acceptor site is a glutamic acid side chain. The polymers are heterogeneous in that they contain varying length and their complexity depends on the levels of branching. PAR chains are highly charged since there are two negatively charged phosphate moieties between ribose sugar groups. These highly anionic polymers are thought to compete with DNA for the binding of chromatin-associated proteins. C.) The Poly(ADP-ribose) cycle. PARP-1 binds and condenses chromatin until DNA-damage induces its activation. Modified PARP-1 and other chromatin associated proteins are then released from DNA. Naked DNA is accessible by DNA-repair and transcription machinery. Meanwhile the poly-ADP-ribose polymers are removed by Poly(ADP-ribose) Glycohydrolase (PARG). After PAR removal, PARP-1 and other chromatin associated proteins can return to their architectural roles.

Within a decade of the discovery of the poly-ADP-ribose polymers, it was observed that the polymers were subject to degradation. In fact, a whole new class of enzyme responsible for the poly-ADP-ribose catabolism was discovered in that context. The enzyme exhibited a glyco-hydrolase activity and was aptly named Poly(ADP-ribose) Glycohydrolase (PARG) [9].

Almost immediately after poly-ADP-ribose polymers were characterized, they were found covalently linked to nuclear proteins (**Figure 1.1B**). The discovery of ADP-ribose linkages to nuclear proteins revealed a new purpose for the polymers as covalent protein modifications [10]. It was shown that poly-ADP-ribose was associated with chromatin and specifically enriched in histone fractions. Following the discovery of the poly-ADP-ribose histone link the topic has been and remains of great interest to the field of chromatin and DNA repair alike. Previously the identification of the particular histone targets or more specifically the amino acid residues that contained the covalent linkage were hard to decipher. Recent advances in analytical analysis techniques like mass spectrometry have allowed for the precise mapping of the ADP-ribose modification sites on the histones [11]. According to peptide mapping experiments, the major modification sites appear not to be the surface exposed glutamic acids as previously suspected. Rather, the target sites reside in the flexible tail regions of the histones.

Chemically, the poly-ADP-ribose polymers are quite similar to nucleic acid. Since a large concentration of the poly-ADP-ribose has been observed in the nucleus, it is reasonable to consider the possibility that poly-ADP-ribose might serve as an alternative substrate for DNA binding proteins. Not surprisingly, the acidic polymers were indeed shown to serve as binding platforms for basic histone proteins [12]. Early studies on how histones interact with poly-ADP-ribose were quite revealing. The histones could bind the ADP-ribose polymers with a very high affinity, which varied depending on the branched nature of the polymers. The affinities were so strong that neither chaotropic salts nor strong acids had any effect in ablating the histone-ADP-ribose interaction. Intriguingly, the interaction could be disrupted if DNA, the preferred histone substrate, was present [12]. Immediately following the discovery of non-covalent histone-ADP-ribose interaction, a new role for the poly-ADP-ribose was described. It was hypothesized that DNA-damage induced the formation of poly-ADP-ribose polymers, which in turn

served as a temporary binding platform for histones while the DNA is repaired. The interaction was shown to be temporary because the polymers were immediately degraded, presumably after the histones rebound DNA, by Poly (ADP-ribose) Glycohydrolase (PARG). The emerging view from that study was that poly-ADP-ribose served as a histone shuttle for a brief period while the DNA damage that resulted in its formation was repaired [13]. In other words, poly-ADP-ribose was shown to be involved in a DNA-damage induced cycle: DNA damage triggers the formation of the polymer, the polymer serves as a histone sink near the damage site, the DNA is repaired rapidly in the histone-less “naked” state and the ADP-ribose is then rapidly degraded allowing the histones to return to the newly repaired DNA substrate (**Figure 1.1C**).

1.2 The Poly(ADP-ribose) Polymerase Superfamily

Meanwhile, many other studies focused on the enzymatic source of the poly-ADP-ribose polymers, the polymerase itself. The enzyme responsible for the synthesis of the ADP-ribose polymers was named Poly(ADP-ribose) Polymerase (PARP) and was occasionally referred to by an older name, Poly(ADPR) synthetase (used as far back as the nineteen seventies) [5]. According to the National Center for Biotechnology Information (NCBI), the enzyme is known by several aliases, the most frequently used ADP-Ribosyl Transferase (ADPRT). For nearly 30 years the PARP enzyme was believed to be the sole source of ADP-ribose polymerization. It wasn't until 1995 that a plant homologue of mammalian PARP was discovered in the plant species *Arabidopsis thaliana* [14]. Later it was shown that two independent PARP proteins were present in *A. thaliana* [15].

Cell lines derived in a mouse PARP^{-/-} background retained the ability to synthesize poly-ADP-ribose when treated with DNA-damaging agents [16]. A closer investigation of the PARP^{-/-} cell lines lead directly to the discovery of a new mammalian enzyme with poly(ADP-ribose) polymerase activity, the enzyme was named PARP-2 [17]. Later it was shown that when both PARP-1 and PARP-2 are ablated, null mouse embryos do not live past gastrulation [18]. Accordingly, the amount of poly-ADP-ribose

produced in the PARP-1^{-/-} cells was severely diminished. When the two enzymes were compared in vitro, it was shown that PARP-2 could only synthesize a fraction of the levels poly-(ADP-ribose produced by PARP-1 (~18-fold lower) [18]. In terms of cellular survival in the presence of DNA-damaging agents the lower enzymatic activity of PARP-2 seems to be sufficient to cover the PARP-1^{-/-} phenotype.

Preceding the description of PARP-2 was the discovery of another PARP-family member, Tankyrase [19]. Tankyrase has sequence homology with both the enzymatic region of PARP-1 and the ankyrin-repeat regions of the ankyrin class of proteins. The localization of Tankyrase to telomeric DNA and its ability to synthesize polymers of ADP-ribose signified the importance of the polymers in regulating cellular processes other than DNA-damage repair.

While PARP-2 and Tankyrase were being described a simultaneous search for other PARP-like gene products was underway. Later, cDNA cloning techniques and homology comparisons helped define the existence of PARP-2 independently of Ame *et al.* [20]. The communication describing the cDNA cloning technique also described the existence of yet another PARP-family member, PARP-3. The PARP-3 protein was later described as being primarily localized at the centriole. The enzymatic activity of PARP-3 was shown to be DNA-dependent and the PARP-3 derived poly-ADP-ribose polymer is functionally involved in the negative regulation of cell cycle progression [21, 22].

In all, a total of 18 proteins have now been characterized as members of the PARP-family. The requisite for membership is a shared homology with the enzymatic domain of PARP-1 (reviewed in [23, 24]). It is important to note that structural homology to the PARP-enzymatic domain does not necessarily mean that the protein is a functional polymerase. A recent comparison of the enzymatic properties of the PARP-family discussed the need to reclassify the family into 3 subgroups [25]. Although all PARP-family members share homology with PARP-1, certain members lack critical amino acids within the enzymatic domain. A fully functional polymerase would belong to group i.) a PARP-polymerase. Key differences in the active site or the substrate recognition region of the enzymatic domain leads to differences in the chemistry and the overall polymerase activity, and such proteins would be classified to the ii.) mono-ADP-Ribosyl Transferase (mART) group. Finally, a complete lack of activity would define

a PARP-family member as belonging to group iii.) the functionally inactive enzyme category. As a result of the study by Kleine *et al.*, an effort has been made to rename the entire family based on the shared homology of the PARP-family with a more ancient class of proteins, the diphtheria toxin (a mART in its own right) [26]. The three classes of the PARP-family are diverse, and the roles that many of the members perform within the cell have yet to be elucidated.

1.3 Poly (ADP-ribose) polymerase-1, a structural and functional description.

As was discussed previously, the enzymatic product of PARP-1 protein have been intensely investigated over the past several decades. In the late 1970's, several studies focusing on the enzyme itself started to surface. From those reports it was shown that the enzyme was on the order of 110-130 kDa in size. It was also shown that the PARP-1 protein primarily localized within the nuclei of several species of eukaryotes [27, 28]. The work of Yoshihara *et al.* provided for a thorough enzymatic analysis, which was greatly aided by the researcher's ability to purify, to a high degree, the PARP-1 protein from bovine thymus preparations. The enzyme was shown to have a half-maximal enzyme rate (K_m) of $\sim 60\mu\text{M}$ for NAD^+ (see Chapter 2 for comparisons of different allosteric activators, or DNA substrates). For the first time researchers deciphered the DNA-to-enzyme ratio for effective catalytic activation [28]. It was shown that the enzyme required an equivalent mass concentration of DNA. In other words, effective enzymatic activation required the formation of a one-to-one complex of PARP-1 to DNA even in the absence of histones (then thought to be the major activators of the enzyme) [28]. Also from that study, it was shown that the enzyme requires a double-stranded, damaged-DNA substrate with a minimum length of 10 bp, and for effective activation a divalent metal, namely magnesium, must also be present. Finally one of the major nuclear proteins thought to be necessary for enzymatic activation, the histone class of proteins, turned-out to be not required. An added level of complication came from the experiments of the past decades, which showed that the PARP-1 enzyme was responsible for the formation of very

heterogeneous polymers (variation in degree of branching and in length). Taken together, what was emerging was a description of a very complex enzyme.

The complicated nature of the PARP-1 protein has remained a major challenge for researchers over the years. The biochemists' approach of minimizing the number of variables in a given problem no doubt guided some of the ensuing characterizations of the PARP-1 enzyme. Thanks to the purification protocols of Yoshihara *et al.*, large quantities of the pure enzyme were attainable [28]. Once purified protein was in hand, efforts to understand the structure-function relationship of PARP-1 could get underway. A great deal of effort was spent on creating anti-PARP-1 antibodies to better define the types of interactions and/or the binding partners of PARP-1 [29-31]. Some researchers even thought to better define the specific regions of the PARP-1 protein that were critical for said interactions [30]. Limited proteolysis was used to generate monoclonal antibodies for specific PARP-1 fragments. From limited proteolysis, it was shown that the protein was comprised of several functional domains [30, 31]. When proteolytically digested, the 113 kDa enzyme released three major fragments of roughly 22 kDa, 46 kDa and 56 kDa in size. The 22 kDa fragment was identified as the automodification region [30]. Further characterization by DNA-cellulose chromatography revealed that the 46 kDa fragment was responsible for the DNA binding properties of the enzyme. The other major fragment, the 56 kDa piece, contained the NAD⁺ binding activity [30]. Taken together, it appeared that the protein was comprised of three domains, the DNA-binding region, the automodification region and finally the catalytic component.

As a result of a careful analysis of the enzyme's dependence on metal ions, it was shown that PARP-1 required both magnesium and Zn(II) for enzymatic activity to be present [32]. It should be noted that the discoverers of the Zn(II) requirement cautioned against the assumption that the enzymatic consumption of NAD⁺ relied on Zn(II). The elegant experiments of Menissier-de Murcia *et al.*, revealed that the enzyme required two Zn(II) atoms, not for enzymatic catalysis, but for DNA-binding functionality [33]. These researchers defined further the region of DNA binding to a cysteine rich stretch of amino acids in the N-terminal region of PARP-1. The Zn(II) binding area of PARP-1 was described as a tandem repeat of what appeared to be a putative zinc-finger motif [33]. The Zn-finger domains were

further defined as comprising the N-terminal 24kDa of the protein. Several years later, the Riken Structural Genomic and Proteomics group of Japan published the Nuclear Magnetic Resonance (NMR) solution structures of the PARP-1 Zn-finger domains, Zn1 and Zn2 [34]. The PARP-1 Zn-fingers are described as containing a rarely seen Zn-coordinating motif comprised of Cysteine-Histidine-Cysteine-Cysteine pattern. The rarity comes from the long stretch of amino acids that separate the critical zinc-binding residues. Interestingly, the DNA-repair protein, DNA-ligase III α , appears to share homology with the PARP-1 Zn-finger domains [35].

Further structural information has provided atomic-level information on how PARP-1 binds DNA [35, 36]. The sequence identity between the two Zn-fingers is not identical and according to recent publications, the Zn2 domain has a greater affinity for DNA [35, 36]. The crystallographic and solution structures of the PARP-1 Zn-fingers in complex with DNA describe a high-affinity interaction with multiple DNA-damage substrates (small oligonucleotide models of DNA damage). The solution studies of Eustermann *et al.* on a Zn1-Zn2 protein fragment (amino acids 1-209), also known as the DNA-binding domain (DBD), (a fragment similar to the one described by Menissier-de Murcia *et al.*) described a dynamic interaction with DNA. The NMR data of the Zn1-Zn2 protein-DNA complex may even indicate a specific binding orientation [35]. The recent structures describing how PARP-1 interacts with DNA are exceedingly valuable and contribute to the understanding of how the protein interacts with DNA damage. What remains to be elucidated is the interaction of the Zn-fingers, or the entire enzyme, with the non-damaged DNA substrate PARP-1 is known to bind, chromatin.

The domain mapping of PARP-1 by limited proteolysis revealed the major functional regions of the enzyme [30, 31]. However, the structural and biochemical work that defined the DBD revealed a major discrepancy between the apparent 46kDa DNA-binding proteolytic fragment and the 24kDa mass of the tandem Zn-fingers. Later it was shown that during apoptosis, the PARP-1 protein is a major target for proteolytic cleavage upon treatment of cells with chemotherapeutics [37]. It was shown that cellular levels of poly(ADP-ribose) during apoptosis were significantly lower even in the presence of extensive DNA endonuclease activity. The work by Kaufman *et al.* revealed that apoptosis related proteolytic

activity results in the cleavage of PARP into two fragments of roughly 25kDa and an 85kDa in size. It was shown that the 25kDa fragment contained the DNA binding activity, the N-terminal Zn1 and Zn2 domains, and the C-terminal catalytic domain was contained in the 85kDa fragment. Interestingly, the C-terminal cleavage product retained only basal levels of enzymatic activity because the enzyme could no longer be induced by DNA strand breaks. Shortly after the apoptosis link was revealed, a novel caspase family member was described as the protease responsible for the PARP-1 cleavage [38]. The discovery of PARP-1 as a major target during the early stages of apoptosis revealed an important dichotomy in the apparent multiple roles of PARP-1. For instance, PARP-1 was responsible for alerting the cell to DNA damage yet, its enzymatic activity must be silenced in the early stages of apoptosis. If PARP-1 were not silenced, the cellular store of NAD^+ would be depleted very rapidly by the enzyme.

Recently, PARP-1 has been of renewed interest to many different structural biology research groups. The Riken Genomics and Proteomics group of Japan were successful in large-scale recombinant expression of several different PARP-1 domains for NMR studies [34]. The researchers who published the first PARP-1 domain structures, the Zn1 and Zn2 domains, followed up those studies with an additional crystallographic structure. The group also published the solution structures of a domain with homology to the C-terminus of the Breast Cancer type 1 susceptibility protein (BRCA-1) (pdb 2COK). The solution structures were published without any mention of biological function. However, homology to this domain is quite common in DNA repair pathway members and it is known that the interactions between the pathway members X-Ray Cross-Complementing group 1 (XRCC-1), DNA-ligase III and PARP-1 are mediated through interactions between the BRCT-domains of those Base-Excision Repair (BER) members [39]. Currently, there are two independent structures of the BRCT domain of PARP-1 in the Protein Data Bank including a recent submission by Loeffler *et al.* [40]. The authors who published the second BRCT structure discuss an apparent lack of a true interaction with the BRCT domain of the XRCC-1 protein. The authors suggest that the PARP-1 interaction with XRCC-1 might be dependent on the presence of covalently linked ADP-ribose moieties, as depicted in **Figure 1B** [40].

Following the publication of the BRCT domain of PARP-1, the Riken group published the solution structures of a novel domain found on the C-terminal side of the BRCT domain, the WGR domain. The domain was named for a conserved three amino acid motif of W, G and R. The WGR domain has no known function but it is conserved in poly(A) polymerases [41]. The three most active members of the PARP-family, PARP-proteins 1-3, contain WGR domains immediately upstream of the C-terminal catalytic or enzymatic domain. A study by Altmeyer *et al.* showed that the WGR domain was necessary for efficient activation of the enzyme. From a chimera study, it was shown that even though the domains of all three proteins share a high sequence homology, the WGR domains could only effectively activate the protein to which they naturally belong [42].

Arguably the most important domain structure for PARP-1 was solved through X-ray crystallography in 1996 [43]. The structure was of the chicken C-terminal PARP-homology domain, also known as the enzymatic domain. Shortly after the publication of the first catalytic domain structure, the same structural biologists published a number of other structures including; the human catalytic domain structure alone and in complex with several inhibitors and most importantly a structure of the catalytic domain in complex with an NAD^+ analogue, carba- NAD^+ . In the manuscript describing the crystal structure of the catalytic domain complexed with carba- NAD^+ , Ruf *et al.* claimed they were able to determine the mechanism for the poly-ADP-ribose reaction [44]. The enzyme-bound NAD^+ analogue allowed the researchers to hypothesize which amino acids were responsible for NAD^+ catalysis, poly-ADP-ribose elongation, branching. Additionally, the researchers defined the critical amino acid for the polymerase function of the enzyme to be glutamate 988. An enzyme without the critical glutamic acid at position 988 would result in a reaction that catalyzes the transfer of a single ADP-ribose moiety [25]. Additionally, the numerous crystal structures of the catalytic domain in complex with the different classes of inhibitors no doubt allowed for the development of pharmaceutically relevant small molecule inhibitor compounds [45]. Several PARP-inhibitors are now in clinical trials and they appear to be a promising new class of molecules in the treatment of cancer [46].

One of the most recent discoveries in terms of PARP-1 structural domains was of a third Zn(II) binding domain. A solution structure based on NMR spectra and a crystal structure of the domain were published nearly simultaneously [47, 48]. Both groups found that the Zn₃-domain was of a novel fold in that its Zn-coordinating ability was not involved in DNA binding. The crystal structure revealed what appeared to be a dimerization interface between two molecules in the crystal lattice [48]. However, no dimerization was detected in the solution structure [47]. Both groups were able to test the enzymatic function in a PARP-1 construct with a Zn₃ deletion, or in the case of Langelier *et al.*, a critical mutation [47, 48]. Collectively, the structural studies on the Zn₃ domain indicate that the Zinc-ribbon structure is critical for mediating contacts with other PARP-1 domains, which in turn coordinate the DNA-binding event at the N-terminus with the activation of the enzymatic function at the C-terminus of the protein.

Recently Langelier *et al.* solved the crystal structure of a near full-length PARP-1 complex with a DNA damage substrate [49]. From stability studies, the authors speculated that the catalytic domain becomes unstable when the other domains are bound to DNA. The destabilized catalytic domain is thought to undergo major conformational dynamics. With increased motion and internal stability changes, the NAD⁺ binding pocket was thought to undergo major changes that would directly influence the binding of NAD⁺ within the catalytic domain [49]. Additionally, when the other domains were bound to DNA, Langelier *et al.* inferred that they would be stabilized and would better serve as a modification target since only the catalytic domain was destabilized and in turn was more dynamically able to modify non-moving target amino acids. The latest crystal structure and the stability studies are the closest anyone has come to date in determining the mechanism of PARP-1 activation. Given the flexibility of PARP-1, a low-resolution solution study or modern electron microscopy techniques involving the full-length protein, both with and without DNA, might help to shed light on the DNA induced dynamics of the C-terminal catalytic domain.

The modular nature of PARP-1 is highly reminiscent to other nuclear proteins which are either natively unfolded or contain regions of order interrupted by disordered stretches of amino acids (reviewed by Dyson and Dyson and Wright) [50, 51]. Although PARP-1 contains six domains that retain a definite

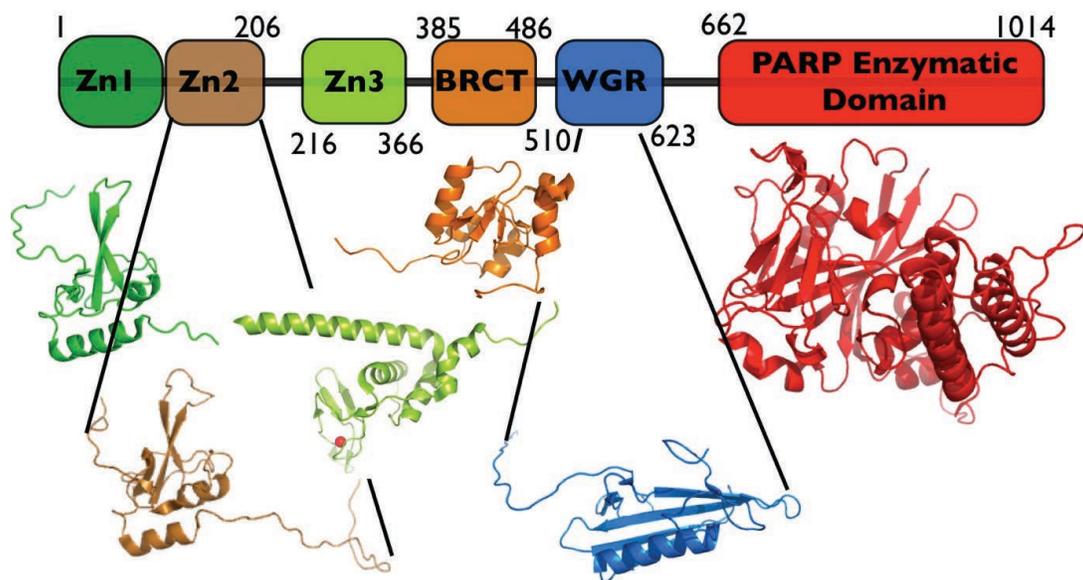


Figure 1.2. The domain organization of Poly(ADP-ribose) Polymerase-1 (PARP-1). PARP-1 is a modular protein consisting of 6 folded structural domains that are connected by flexible linker domains. From left to right the protein is comprised of the DNA binding domain (DBD), which includes the Zinc fingers, Zn1 and Zn2 (residues 1-206). The Zn3 domain is a novel protein fold that both coordinates zinc atom yet it is not involved in DNA binding rather, it is thought to mediate inter-domain interactions. The BRCT domain shares homology with the BRCA-1 C-terminus and it is thought that BRCT serves multiple functions in PARP-1, among them include mediating interactions with other BRCT-domains of known binding partners and serving as the primary target for automodification by the C-terminal PARP-Enzymatic or Catalytic domain. Between the BRCT and the Catalytic domain resides a WGR domain, which is also thought to have a weak DNA binding function. Below the domain map representation are the 3-dimensional structures taken from NMR and crystallographic studies (PDB codes 2DMJ, 2CS2,2RIQ, 2COK, 2CR9 and 2PAW from N- to C-terminus, respectively).

three-dimensional structure, roughly 10% of the protein is made up of flexible linker regions (**Figure 1.2**). The linkers contain highly charged and polar amino acids (rich in Lys, Ala, Glu, Pro, Arg, Ser and Gly residues). These unfolded regions can be thought of as flexible loops that serve to link the independent roles of the multiple domains for optimal enzyme functionality. PARP-1 may have evolved to be flexible in order to properly interact with its known multiple binding partners. It is likely that the enzyme can take on multiple orientations and binding partners so that it can effectively participate in several nuclear pathways. Even some of the strongest biological interactions known, like the interaction of an antibody with its epitope, have been shown to be stronger if the polypeptide for which the antibody recognizes is relatively disordered [52]. A recent comparison of chromatin associated proteins revealed a high degree of disorder amongst remodeler proteins [53]. Proteins that perform their functions in the context of large complexes may have evolved to have regions of disorder as a means to better interact with multiple substrates that might include DNA, histones and other chromatin modifiers. Additionally, the flexibility of PARP-1 might be related to intra- and inter-domain interactions that must occur for optimal enzymatic function. Besides the known ability of PARP-1 to automodify certain domains or regions of itself, there is some evidence that different domains of PARP-1 have the potential to interact with each other both in cis and trans [47, 49].

There may be other reasons why PARP-1 has so many linkers and is such a flexible molecule, which may have to do with enzyme processivity. In fact, several decades ago a thorough biophysical analysis was performed on highly purified bovine PARP-1. In that study, Ohgushi *et al.* discussed the elongated nature of the polypeptide and reasoned that the enzyme must undergo a conformational change upon binding DNA [54]. The researchers hypothesized that DNA binding and catalytic activation were separate events, and that DNA binding occurred prior to the induction of the enzymatic activity. Later, a structural study demonstrated that the catalytic domain of PARP-1 had an altered conformation when an NAD⁺ analog, FR257517 (a competitive inhibitor), was co-crystallized within the enzymatic domain [55]. The biologists who solved the inhibitor-bound PARP-1 structure reasoned that the relatively large conformational change in the catalytic domain was due to the movement of a single amino acid, Arginine

878, an amino acid in the region known as the adenine-ribose binding site. Basically, the binding of the NAD-like molecule to a small region within the PARP-1 catalytic domain induces a large conformational change that could echos throughout the greater domain structure.

1.4 PARP-1 in DNA damage repair

The flexibility and modular make up of PARP-1 allows it to undergo multiple conformations and bind many DNA and protein substrates. In the absence of DNA damage, poly-ADP-ribose is barely detectable [56]. Upon induction of DNA damage, the poly-ADP-ribose polymer concentration increases by as much as 100 to 500-fold, 90% of which is synthesized by PARP-1 [56, 57]. If PARP-1 were overactivated within the cell, in a case where DNA damage levels were very high, the enzyme has been shown to deplete the cellular stores of NAD^+ . In fact it has been shown that over activated PARP-1, if unchecked by caspases, would result in cellular death by necrosis due to severe NAD^+ depletion [58]. In order for PARP-1 to exist without depleting the NAD^+ stores of a cell, the enzyme is thought to be inactive in the absence of DNA damage. Some years back it was shown that PARP-1 is very sensitive to the structure of nicked DNA [33, 59]. It is thought that PARP-1 exists in its inactive form, or enzymatically silent state, until it binds broken or free DNA ends to protect the genome for unchecked DNA damage sites [60]. Among the several DNA structures recognized by PARP-1, a nick represents one of the preferred binding substrates over that of a double stranded break DNA. A study with a 139bp double-stranded DNA indicated that a centrally located nick results in a V-shaped structure [61]. From electron micrographs, it was shown that PARP-1 preferentially bound the apex of the V-shaped nicked DNA, over the double-strand DNA ends. A recent structural study by Eustermann *et al.* showed that PARP1 had an increased in affinity for a gapped DNA over a nicked version of the same sequence, indicating to the authors that flexibility matters in terms of binding affinity [62]. The preference for single-strand over double-strand breaks (DSBs) helps explain why PARP-1 is so critically connected to the base-excision repair (BER) pathway, as well as to single-strand break repair (SSBR) [63, 64].

It is known that the PARP-1 enzyme has an extremely high specific activity and that it primarily targets itself through automodification [28, 54, 65]. Automodification of PARP-1 serves as a signal to the DNA repair machinery that damage is present. The ADP-ribose polymers that remain anchored to the PARP-1 protein are thought to bring the base-excision repair protein complex to vicinity of the site of DNA damage. The BRCT-domain of PARP-1 has homology to similar domains present in many of the XRCC-1 BER and SSBR protein complexes [39, 63, 66-68]. It is thought that the BRCT-homology domains serve as key sites for protein interactions. However, a careful biophysical study that investigated the BRCA-domain binding properties of a number of these related domains found that domain interactions may rely on their modification by ADP-ribose polymers [40]. Regardless of how the proteins interact, it is important to note that the critical event for recruiting the repair machinery is the formation of ADP-ribose polymers after the PARP-1 DNA-binding occurs. The recognition and binding of damaged DNA by PARP-1 seems to be a critical factor in both the BER and SSBR repair pathways. Subsequent recruitment of the XRCC-1 complex allows for the binding of PARP-1 by complex members, most likely DNA-ligase III and XRCC-1 proteins. Once bound to XRCC-1, the enzymatic activity of PARP-1 drastically diminishes, there-by conserving NAD^+ [68]. Additionally, the PARP-1/XRCC-1/DNA Ligase III complex seems to be quite stable and may even serve to function in an alternative or redundant DSB repair pathway [66].

The involvement of PARP-1 in several different DNA damaged repair pathways has lead some to regard the enzyme as “the guardian angel” of the genome [69]. For example, comparative analysis of PARP-1 and another important DNA repair complex, DNA dependent Protein Kinase (DNA-PK), indicates that PARP-1 has a notably higher affinity for several types of DNA damage [70]. The interplay between PARP-1 and DNA-PK was first addressed in a key genetic study in the 1990s [71]. Another group has shown that PARP-1 may cooperate with the heterodimer Ku70/Ku80 to increase the affinity of binding of the possible ternary PARP-1/Ku70/Ku80 complex to matrix attachment regions (MARs) of DNA flanking the coding region of immunoglobulin μ heavy chains [72]. The MAR regions are thought to have increased flexibility. Although flexible DNA does not mimic a site of DNA damage per se, the

hypothesis is that increased flexibility in DNA increases PARP1's ability to bind. The binding of non-damaged bent DNA like that of MARs, might serve to sequester PARP-1 so that it can participate in V(D)J recombination. V(D)J recombination is critical for the maintenance of a functional immune system and represents large scale events that may appear to PARP-1 as DNA damage. In, a study by Brown *et al.*, PARP-1 and specifically its ADP-ribose polymers, were critical for activating the DNA-damage pathway if the DNA-PK catalytic subunit were absent, a situation that results in an immune-deficient phenotype [73]. Brown *et al.* reasoned that in severe combined immune deficient (scid) cells, PARP-1 is an obstacle to effective V(D)J recombination because it binds DNA ends too effectively. However, upon automodification, PAR-bound PARP-1 loses affinity for DNA and can be replaced by the DNA-PK complex, in turn allowing for effective DNA repair, but not effective V(D)J recombination [73].

Another study investigated V(D)J recombination and found that, as might be expected, PARP1 automodification was present at DNA ends. The resulting poly-ADP-ribose was shown to be important during the PARP-mediated recruitment of DNA-PK. Additionally, it was shown that PARP-1 is involved in deciding the type of DNA repair that will result once the DNA-PK is recruited [74]. Paddock *et al.* discussed important interplay between DNA-PK and PARP-1 in their recent study. In that study, the authors demonstrated binding competition between PARP-1 and Ku70, a component of the DNA-PK complex, which controls the type of DNA repair the cell undergoes; high-fidelity or mutagenic repair [74]. Recently, the interaction between PARP-1 and DNA-PK was observed in single molecule electron microscopy studies [75]. The complex indicates that PARP-1 may help to alter DNA-PK conformations on DNA. The evidence for this claim comes in the form of a small-angle x-ray scattering (SAXS) solution model of the DNA-PK molecule without PARP-1 present [76]. Comparing the electron micrograph derived model with the SAXS model results in two significantly different conformations of the DNA-PK complex. Spagnolo *et al.* speculate that PARP-1 may influence the structure of the DNA-PK molecule so that the enzyme complex can undergo auto-phosphorylation [75].

In the sense that PARP-1 binds and protects ends from recombination, PARP-1 may serve as an anti-recombinogenic factor [69]. According to Tong *et al.* this particular property of PARP-1 is important for preventing unwanted and deleterious chromosomal recombination events. Additionally, PARP-1 null murine cells display increased levels of aneuploidy, chromosomal fragmentation, fusion and loss, some of which result from telomeric dis-regulation [77-79]. The decades of work linking PARP-1 to the DNA damage response has made it an obvious target for pharmaceutical inhibition [46, 80]. Recent studies have shown that PARP-1 inhibition may be especially effective in killing cells when combined with chemotherapeutics, which are deficient in performing DNA repair through the Homologous Recombination (HR) pathway; PARP inhibition seems to be especially promising as an anti-cancer agent in cells deficient in BRCA-1 and BRCA-2 proteins [81-83]. The mode of action that PARP-1 inhibitors have on HR-deficient cells seems to be a method to further sensitize cells to SSBs. PARP-1 inhibition results in an increase in the number of DNA lesions that would normally be repaired through the PARP-1-mediated SSBR or BER pathways. In cells deficient in HR, like cells with *BRCA-1* or *BRCA-2* deletions or mutations, unrepaired single-strand breaks become unrepaired double-strand breaks that result in chromatid breaks and further aberrations. The phenotype is thought to be so severe that the cells would cease to be viable if they were further assaulted through PARP inhibition [81].

The involvement of PARP-1 in DNA repair seems to be ubiquitous. PARP-1 has demonstrated roles in several types of DNA damage repair including single-strand break repair (SSBR), base excision repair (BER), non-homologous end joining (NHEJ) (through its interactions with DNA-PK), and is indirectly involved in homologous recombination (HR). Many studies link the enzymatic activity of PARP-1 to its roles in DNA repair. It is through its involvement in DNA repair that PARP-1 is thought to protect or “guard” the genome. In fact, the work of Beneke, Burkle and colleagues has linked PARP-1 function to longevity in mammals [84-90]. Their work has led to the notion that the enzymatic activity of PARP-1 varies among species and that variation correlates positively with lifespan (species with higher PARP-1 specific activity maintain longer lifespans) [91]. Beneke and Burkle cite subtle differences in the primary structure of PARP-1 and these differences represent an evolutionary “control” [92]. In that work

Human PARP-1 had a relative activity level roughly 2-fold above that of the Rat PARP-1. Interestingly, certain single nucleotide polymorphisms (SNPs) within the enzyme's active site have been documented within the human population [93, 94]. It is known that these SNPs result in increased cancer susceptibility [95, 96] and have been shown to have major enzymatic implications *in vitro* [93, 94]. According to Beneke and Burkle, the major differences between species might depend less on the active site of the enzyme and more to sites of automodification, other post-translational modifications, and/or regions critical for optimal protein-protein interactions [92].

It is known that the DNA binding of PARP-1 is the crucial initiating step for the enzyme [33, 59]. Additionally, several researchers have indicated specific domains of PARP-1 are responsible for mediating the protein-protein interactions between PARP-1 and the other members of the DNA repair pathways [39, 63, 66-68]. What is not clear from the research to date is how the domains of PARP-1 interact in the presence or absence of DNA damage. To better understand the fluid nature of the PARP-1 interaction with other key DNA repair proteins, a clear understanding of how the protein behaves alone is required. Several investigations have described the binding of different DNA structures by PARP-1. However, side-by-side comparisons of the different DNA-damage models (i.e. single-strand breaks (nicks), overhang and blunt-ended DNA damage models) have yet to be performed. A better description of how PARP-1 binds different DNA structures is needed. It is thought that PAR-bound PARP-1 loses its ability to bind DNA due to charge repulsion of the long polymers of ADP-ribose [92]. PARP-1 has been shown to bind DNA at its N-terminal DNA-binding domain, specifically its Zn1 and Zn2 domains [33, 59]. Altmeyer *et al.* and Tao *et al.* independently showed that PARP-1 primarily automodifies itself in the central region of the protein, relatively far from the DNA binding domain [42, 97]. The relative impact of how the automodification on the central region of PARP-1 affects the DNA binding at the N-terminus of the protein remains to be elucidated. The major question that arises is how that relates to the flexible nature of the protein. Could the flexible and independent Zn-fingers [35] retain their DNA binding ability in the presence of large negatively charged polymers located some distance downstream? In addition to determining thermodynamic differences between different DNA structures, a thorough

comparison of the enzymatic parameters of PARP-1 should be undertaken. For instance is the enzymatic efficiency (k_{cat}/K_M) of PARP-1 somehow dependent on the structure of its allosteric activator, DNA? Addressing such questions might lead to improved cancer therapies in the future, since different chemotherapeutics lead to different types of DNA damage (see review [98]). PARP-1 was recently linked to a the DNA-damage response of platinum-based chemotherapeutics [99]. Since all PARP-family members share homology in the catalytic domain, competitive inhibitors that target that domain are likely to interfere with non-DNA damage related pathways. The use of PARP inhibitors may be enhanced if combined with specific chemotherapeutics that are known to cause the type of damage PARP-1 recognizes. By eliminating some of the redundancy in DNA repair pathways, PARP inhibition might serve as a means to better sensitize and kill cancer cells.

1.5 PARP-1 in Chromatin and Transcription

PARP-1 is one of the most abundant proteins within the nucleus. It is estimated that between 10^5 and 10^6 molecules of PARP-1 are present in the nucleus at any given time [100, 101]. It is no wonder that PARP-1 has been implicated in nearly every type of DNA damage repair. With nearly a million molecules constantly scanning for DNA damage, one might expect sites of DNA damage to be discovered quite quickly. Another effect of having so many enzymes present in the nucleus could be that PARP-1 may inadvertently contaminate nuclear extracts and complicate other types of experiments. In fact, during the early 1980s the research of Matsui *et al.* implicated a nuclear factor necessary for effective RNA Polymerase initiation [102]. The protein or factor implicated was named TF_{II}C but was later shown to be PARP-1 (reviewed in [103]). The connection between PARP-1 and transcription by RNA Polymerase II became evident a short time later through the work of Hough and Smulson [104], where it was shown that PARP-1 was not a contaminant in the transcription experiment after all. Hough and Smulson made that connection that poly-ADP-ribose strands were indeed involved in transcription [104]. The authors showed that ADP-ribosylation of the base unit of chromatin, the histones that comprise the

nucleosome [105], resulted in chromatin that was more sensitive to nuclease activity. The implication was that the enzymatic activity of PARP-1 was associated with histone modification, which resulted in a more open conformation of DNA, and an increase in RNA Polymerase II transcription [104]. However, that same study showed poly-ADP-ribose was not an absolute requirement for active transcription for all genes. The role of PARP-1 in transcription remained unclear until the discovery of ADP-ribose bound histones. The experiments of Loetscher *et al.* implicated the product of PARP-1 in a signaling pathway [106]. The authors knew that cellular NAD^+ levels were related to PARP-1 activity. In their study, monolayers of hepatocytes were incubated with media containing increasing amounts of NAD^+ and as a result, increased chromatin-associated poly-ADP-ribose was observed. The authors inferred that cellular NAD^+ levels represent a signal to chromatin by PARP-1. In other words, PARP-1 is an important signaling molecule for chromatin and it serves to alert the cell of important environmental changes in the cellular surroundings [106]. Other studies have shown that complexities of the poly-ADP-ribose polymers, in terms of branching and linear length, are directly dependent on NAD^+ concentrations within the reaction conditions [107, 108]. Also dependent on the concentrations of NAD^+ used in the reaction, were the levels of poly-ADP-ribose histone modifications [107]. PARP-1 was shown to be the primary target for poly-ADP-ribose modification, and histones were shown to be far less frequent targets for trans-ADP-ribosylation [107, 108].

Regardless of the total level of modification present, PARP-1 was shown to be involved in the modification of histones preceding the efficient initiation of transcription [104]. A major factor that remained unclear at the time was how PARP-1 was able to associate with the more condensed chromatin structure prior to transcription initiation. In 2003 the work of Tulin and Spradling changed our understanding of how PARP-1 functions in transcription [109]. The authors elegantly described the chromatin structure pre- and post-transcription initiation using modern fluorescence microscopy techniques [109]. Furthermore, it was shown that PARP-1 was activated in a DNA damage-independent manner at the promoter region of the gene that codes for the Hsp70 chaperone protein. The authors of that study were able to visualize what they termed “puffs” in their studies of *drosophila* salivary gland

polytene chromatin [109]. The “puffs” were described as short-lived localized PARP-1 activation events whereby PARP-1, histones and transcription factors were covalently modified with ADP-ribose polymers. The ADP-ribose polymers were quickly removed through the enzymatic response of Poly(ADP-ribose) Glycohydrolase (PARG). After poly-ADP-ribose removal, the chromatin once again returned to the condensed appearance with a similar structure to that of the pre-induced polytene chromatin. Interestingly, PARP1 remained present at the chromatin “puffs” throughout the activation event. This is contrary to the common belief that once PARP-1 is modified the negatively charged ADP-ribose polymers repel the like charged DNA [103, 110-112]. It is possible that PARP-1 retained some of its binding affinity to the open euchromatin and this could explain why the molecule was not excluded from the “puffed” regions by charge repulsion. An alternative explanation might be that the large polymers serve to sterically trap PARP-1 in the “puff” region by preventing it from diffusing away from the active transcription site. In an effort to rule out other factors, that may be responsible for the “puffing”, the authors targeted the enzymatic properties of PARP-1. Following the treatment of *drosophila* larvae with 3-aminobenzamidine (3-AB), a PARP inhibitor, the “puff” phenomenon was no longer visible. The lack of “puff” formation and the marked reduction in mRNA transcripts in the presence of PARP-specific inhibitors is highly suggestive that PARP-1 is critically linked to transcription, at least at Hsp70 promoters in *Drosophila*.

The elegant description of PARP-1 induced “puffing” in *drosophila* may help to clarify a separate issue of puffing-related phenomena seen prior to the publication by Tulin and Spradling [109]. Cartwright and Elgin showed that after heat shock induction, the structure of chromatin, the continuous repeated spacing of mononucleosomes, is disrupted at *HSP70* promoters [113]. Interestingly, the authors presumed that the spacing was due to major changes in chromatin structure. PARP-1 activation and the subsequent modification of proximal histones may in fact underlie the cause of the changes in nucleosome spacing in the sense that modified histones would lose their ability to bind DNA effectively and this would result in the decondensation of chromatin. A mechanism consistent with the experimental findings of Tulin and Spradling was described by Cartwright *et al.*, where the expected increased

accessibility by the nuclease DNase I was possible due to the more open structure of the heterochromatin [113].

The importance of PARP-1 in transcription is still being investigated. The implication that PARP-1 is associated with undamaged chromatin prior to transcription initiation is rather intriguing. For instance, how is an enzyme, which recognizes sites of damage activated in the absence of said damage? Two separate groups shed light on this question [114, 115]. Both Kauppinen *et al.*, and Cohen-Arman *et al.*, discovered a mode by which the enzymatic activity of PARP-1 could be enhanced by other cellular enzymes [114, 115]. From studies of neuronal cell cultures the researchers made the connection that PARP-1 mediated cell death could be attenuated if key small molecule kinase inhibitors were added to cell cultures. The kinase inhibitors were specific to the extracellular signal-regulating kinases 1 and 2 (ERK1/2), which belong to the mitogen-activated protein kinase (MAPK) family [116]. It seems that the ERK1/2 kinases are responsible for increasing PARP-1 to a “maximal” level of enzymatic activity in neurons exposed to alkylating agents or in the response of oxidative stress. Kauppinen *et al.*, went further and identified the ERK2-PARP-1 relationship as involving a direct interaction between PARP-1 and the kinases. The authors of that study showed that PARP-1 could be directly phosphorylated on two amino acid residues (residues S372 and T373) [115]. Deletion mutants of the residues proved detrimental to the interaction and resulted in a decrease of PAR synthesis by PARP-1. Additionally, constitutively “on” mutations, in the form of S372E and T373E, resulted in higher PARP-1 activity without ERK2 present. The critical phosphorylation targets for the ERK1/2 kinase family reside in the flexible linker region that joins the Zn3 and BRCT domains [115] (linker region is depicted in Figure 1.2). Others have described the phosphorylation of PARP-1 in different contexts [117, 118]. However, the fact that other enzymes can influence the enzymatic properties of PARP-1 signifies the important roles PARP-1 plays in non-DNA-damage signal transduction pathways. PARP-1 is the protein usually associated with its role as a first-responder in terms of signaling DNA-damage, but increasingly the enzymatic functions of this dynamic protein can be implicated elsewhere. The existence of pathways that are independent of DNA-damage may yet prove critical in understanding the roles of PARP-1 in transcription. Could other post-

transcriptional modifications of PARP-1 serve as prerequisites for transcription initiation at key gene promoters like *HSP70* for instance? The answer to that question remains to be found. However a recent study by Kotovo *et al.*, discussed an indirect phosphorylation event that resulted in the activation of the *drosophila* homologue of PARP-1. In that study the authors discovered a connection between the phosphorylation of H2Av, the homologue to the mammalian histone H2AX, and a resulting increase in PARP-1 activity [119]. The connection between PARP-1 activation and chromatin de-condensation or transcription initiation seems to reflect our expanding knowledge of the sensitivity of PARP-1 to changing environments both intra- and extracellularly. PARP-1 plays an important role in relaying environmental changes, whether it becomes activated by binding DNA-damage or through signal transduction mediated pathways.

The Role of PARP-1 in Chromatin Architecture

As has been discussed, most cells have a large quantity of PARP-1 molecules within the nucleus. Additionally, PARP-1 binds all types of DNA damage models. In the absence of DNA-damage or active transcription, how is such an important molecule kept occupied/sequestered until it is needed for the enzymatic signaling functions. Does PARP-1 bind to undamaged DNA and if so, is it as enzymatically active? Among the researchers who tried to address this persistent question was Gradwohl *et al.* in the late 1980s [120]. In that work, the researchers carefully prepared plasmid DNA in several different forms that included; linearized, nicked, topoisomerase relaxed and closed and supercoiled. Next, purified PARP-1 was incubated with the different forms of DNA and the relative enzymatic activity compared. Finally, the researchers used electron micrographs to determine the types of PARP-bound DNA structures that were present under those conditions [120]. Interestingly, PARP-1 seemed to preferentially bind the supercoiled plasmid DNA even though the enzyme was the least activated in that context. The micrographs, snapshots help describe the types of DNA structures PARP-1 can binding, from those

images it a definite preference for binding to DNA crossovers. In micrograph after micrograph, PARP-1 was found at the center of the x-shaped DNA crossover junctions [120].

Intriguingly, a similar pattern for the binding of superhelical crossovers was seen for the linker histone protein H1 [121, 122]. The H1 protein plays an important role in chromatin compaction and it most likely does this by binding the two duplex strands of DNA at the entry and exit points of the nucleosome. In naked superhelical DNA, the crossovers appear to be the region in which H1 is optimally positioned. The cross-over region represents a position in which H1 can simultaneously bind two strands of DNA [122]. Kim *et al.* showed that H1 and PARP-1 compete for binding at the dyad axis site of nucleosomal arrays [123]. From that work, the authors were able to determine a great deal of information on the H1-PARP-1 interplay on chromatin: 1) only one PARP-1 or H1 molecule binds per dyad axis (both bind 2 duplex strands of DNA monomerically); 2) in micrococcal nuclease (MNase) digestion of chromatin arrays both proteins protect a similar number of additional base pairs of DNA (~160bp and ~165bp for PARP-1 and H1, respectively); 3) like H1, incubating chromatin with PARP-1 decreases RNA polymerase II (Pol II) dependent transcription. Importantly, one major difference between PARP-1 and H1 was established. The addition of NAD⁺ to the PARP-bound chromatin reverses the structural effects of PARP-1, but not of H1. In other words, simply adding the enzyme's catalytic substrate was enough to de-condense chromatin. The de-condensation was shown to be a result of the automodification of PARP-1 and not the covalent addition of PAR moieties to the chromatin, no signs of histone modification were detected [123]. The findings presented in the work by Kim *et al.* suggest an example of a kind dichotomy in protein function [120]. For instance PARP-1 can bind and condense the structure of chromatin and in essence act as a transcription repressor, however, upon activation of its enzymatic domain the repressive activity is reversed to a degree that the role of PARP-1 is reminiscent of an enhancer of RNA transcription.

The work of Kim *et al.* was a major contribution to the understanding of how PARP-1 imparts structural changes on chromatin. What was not elucidated in those important studies was the mechanism by which PARP-1 binds chromatin. For example, the H1 protein is thought to bind chromatin at the dyad

axis of nucleosomes yet, it is only a fraction of the size of PARP-1 (~20kDa and ~110kDa, respectively). Additionally, from MNase digestions of chromatin PARP-1, it was shown to protect less of the linker DNA between nucleosomes than the H1 [123]. Structurally, this is likely since the Zn1-Zn2 comprised DNA binding domain is relatively small (~20% of the protein). Since it is not contributing to the DNase1 protection, where could the other ~80% of the protein's bulk reside? Could the non-DNA binding region of PARP-1 be interacting with the DNA that wraps around the histone octamer? If that were the case, the additional bulk of PARP-1 enzyme would serve as redundant protection to the ~150bp of DNA already protected by the histones themselves. Alternatively the model might be that the additional non-DNA binding regions of PARP-1 would directly interact with the histones surfaces not occluded by the DNA gyres. Support for the latter model came in the form of a 2007 manuscript by Pinnola *et al.* [110]. Pinnola *et al.* showed that the binding of key regions of the histones themselves could influence the enzymatic activity of *Drosophila* PARP-1. Those experiments proved that, at least for *drosophila*, PARP-1 could directly interact and be regulated by histones. Others had shown previously that PARP-1 could interact with histones but the mechanism for interaction was not clearly defined in those reports (reviewed in [103]).

The structural complexity of PARP-1 with its DNA binding region, its automodification region and its catalytic domain, make describing specific interactions, with any measure of detail, very difficult. However, if one were to investigate a specific interaction by using truncation mutants of PARP-1, detail in describing that interaction might result. Thanks to the creation of certain protein truncation constructs, Wacker *et al.* was able to add greater detail to the understanding of the PARP-nucleosome interaction [124]. The authors showed that the N-terminal DNA binding region of PARP-1 binds chromatin with similar affinity as full-length PARP-1. However, without the NAD⁺ binding portion of the C-terminal catalytic domain present, the PARP-1 truncation constructs fail at condensing chromatin. Interestingly, the C-terminus has never been implicated with any function other than enzymatic activity prior to the publication of that study [124]. The critical interactions with chromatin seem to specifically involve the binding of DNA by the N-terminal Zn-fingers. Also described in the Wacker *et al.* study was a construct

that consisted of a full-length PARP-1 with critical Zn-finger mutations that could neither bind DNA nor condense chromatin. These findings suggest that without proper interaction with DNA, the other regions of the protein, like the C-terminal catalytic domain, cannot properly interact with a nucleosome, or nucleosomes, and contribute to chromatin compaction. One question that remains unanswered in terms of the PARP-chromatin interaction is; does Zn-finger binding to non-damaged DNA in the context of chromatin result in a conformational change similar to the change that is hypothesized for the enzyme in the context of DNA damage? According to studies of non-damaged intact plasmid DNA, PARP-1 readily associates with intact circular DNA, but its relative enzymatic activity levels are far decreased when compared to the binding of plasmids that contain nicked regions [121]. Studies of the PARP-1 interactions with intact plasmids suggest that the protein preferentially binds two duplexes of DNA simultaneously [121]. The crossover structures strongly resemble DNA at the entry and exit sites of the nucleosome. Another question that remains is where exactly is PARP-1 binding to the nucleosome could it bind the DNA at the crossover, or at the entry-exit site? According to Kim *et al.*, the PARP-1 protein protects ~13 bps of linker DNA that protrudes from the nucleosome [123]. Is the protection due to the Zn-fingers binding that linker DNA or due to the non-DNA-binding regions of PARP-1 sterically blocking MNase from accessing that stretch of DNA? In an effort to rectify the apparent differences in the NMR and crystallographic structures of Zn₃, Langelier *et al.* found that mutating certain residues within the crystallographic “dimerization” interface of Zn₃ reduced the ability of PARP-1 to condense chromatin [125]. It was from that study that the authors suggested a special case of PARP-1 dimerization as occurring in the context of chromatin but not necessary for its role as a DNA damage sensor. Shortly after this model was proposed however, the same group of researchers altered the model slightly. In that more recent publication by Lanelier *et al.* it was proposed that PARP-1 can more easily incorporate in the bent DNA, like the DNA located near the dyad axis of the nucleosome [36]. Taken together, there has been a lot of important information gathered regarding how PARP-1 can interact with chromatin yet many questions remain.

1.6 Specific Aims

A great many researchers have dedicated the focus of their research toward advancing our understanding of the multiple roles and functions that PARP-1 plays within the cell. We now have a better understanding of the key types of interactions that link PARP-1 to, chromatin dynamics, transcription, aging, and DNA damage repair. Additionally, several PARP-specific small molecule inhibitors are showing promise in clinically in the treatment of patients with breast cancer [126]. Since the small-molecule inhibitors are specific to the NAD⁺ binding region in the catalytic domain of PARP-1, their use could inhibit other non-PARP-1 related cellular functions as well, due to the sequence homology between PARP-family members occurs in the catalytic domain. It is also not known whether there are any long-term side effects of PARP-family inhibition since such studies have yet to be completed. One possible way to avoid unwanted PARP-1 inhibitor side effects might be to better define the specific differences between the diverse roles of PARP-1. Dissecting the specific interactions needed for DNA damage repair and comparing them to the interactions needed for chromatin dynamics may result in an increased drug efficacy for patients who are undergoing treatment with PARP-inhibitors. Increased appreciation for how PARP-1 functions in seemingly opposing cellular functions may lead to a different class of PARP-1 inhibitors entirely.

It is my goal to contribute to more clearly defining how PARP-1 functions to bind *in vitro* models of known DNA-damage and chromatin substrates. Below are the specific aims that have guided the research in this thesis in order to help answer some of the remaining questions related to the biochemistry of PARP-1. The questions addressed here are 1.) How does PARP-1 recognize different types of DNA structures? 2.) Are there differences in the binding modes of PARP-1 to chromatin substrates as opposed to DNA-damage? 3.) Does DNA structure relate to differences in enzymatic activity levels? 4.) How does the catalytic domain become activated upon DNA binding? I hypothesize that the type of DNA substrate, either DNA damage or nucleosomes, recognized by the N-terminus of PARP-1 serve to relay a message to the C-terminal catalytic domain and the resulting level of enzymatic activity is a function of

the distinct molecular conformational rearrangement that is specific to that binding substrate. Additionally, in order to describe the substrate binding at the N-terminus, an intact and fully functional enzyme should be the primary construct used in *in vitro* experiments since truncation mutants by definition can not provide a complete understanding.

Specific Aim 1.) Using biochemical and biophysical techniques including; eukaryotic cell culture, enzymatic activity assays, thermodynamic measurements of binding constants, biophysical studies of PARP-1 bound to its DNA-damage and chromatin substrates, determine relative affinities and enzymatic parameters for the two opposing types of nuclear substrates. Using the above mentioned techniques, the evidence gathered will help to test the hypothesis that the N-terminal DBD can differentiate between substrates and upon binding, those differences may result in varying levels of enzymatic activity. Chapter 2 of this thesis addresses the germane questions regarding the means by which PARP-1 binds and responds enzymatically to unrelated nuclear structures, DNA-damage and chromatin.

Specific Aim 2.) Using biochemical, structural biology and biophysical techniques, elucidate the mode of binding of PARP-1 to double stranded DNA-damage models. Low-resolution small-angle scattering (SAS) techniques including light scattering, small-angle x-ray and neutron scattering (SAXS and SANS) will be used to gather structural information for the average conformer of PARP-1 alone and PARP-DNA complexes in solution. The biophysical techniques combined with molecular Monte Carlo simulations, will be used to test the hypothesis that DNA substrate binding results in a conformational rearrangement throughout the enzyme. Chapters 3, 4 and 5 are dedicated to the low-resolution scattering data collected on two PARP-1 proteins and their complexes with DNA. Chapter 3 will address the DNA binding modes of several DNA-damage models to the N-terminal half of PARP-1. Chapter 4 will address the efforts implored to determine solution structures of full-length PARP-1 and the above-mentioned N-terminal half of PARP-1 from small angle x-ray scattering (SAXS). Chapter 5 will compare and contrast small-angle neutron scattering (SANS) and SAXS data collected on the full-length and N-terminal PARP-1 proteins

alone and in complex with DNA damage substrates. Taken together, the small angle scattering data and computational methods used to model the solution states of the molecules will help describe the DNA-dependent mechanism for enzymatic activation of PARP-1.

Chapter 2

Alternative binding modes of Poly (ADP-ribose) Polymerase 1 to free DNA and nucleosomes

Chapter Overview

The current chapter is a now a published, peer reviewed article in the Journal of Biological Chemistry, which was accepted on July 31st, 2012. The first three authors contributed equally to the publication of this work. The expertise of Dr. Uma Muthurajan and her knowledge of nucleosome assembly allowed for the study of the highly defined chromatin substrates described here. The work and dedication provided by Michael Kramer MS allowed for the quantitative analysis of thermodynamic binding constants and he helped to develop the HI-FI FRET assay needed to study the PARP-1 protein as well as other systems in the laboratory. This author contributed by producing high-quality full-length human PARP-1. Without the full-length PARP-1 protein, the enzymatic properties and high-quality biophysical studies of PARP-1 in complex with its biological substrates would not have been possible.

Summary

Poly (ADP-ribose) Polymerase 1 (PARP-1) is an abundant nuclear protein that binds chromatin and catalyzes the transfer of ADP-ribose groups to itself and to numerous target proteins upon interacting with damaged DNA. The molecular basis for the dual role of PARP-1 as a chromatin architectural protein and a first responder in DNA repair pathways remain unclear. Here we quantify the interactions of full length PARP-1, as well as its N-terminal half with different types of DNA damage and with defined nucleosome substrates. We find that full length PARP-1 prefers nucleosomes with two linker DNA extensions over any other substrate (including several free DNA models), and that the C-terminal half of PARP-1 is necessary for this selectivity. We also assess the ability of various substrates to activate PARP-

1 activity, and find that the most important feature for activation is one free DNA end, rather than tight interaction with the activating nucleic acid. Our data provides insight into the different modes of interaction of this multi-domain protein with nucleosomes and free DNA.

Introduction

PARP-1 is a conserved multi-domain enzyme that is present in all eukaryotes except yeast. With an estimated abundance of $\sim 10^6$ molecules per cell, there is approximately one PARP-1 molecule per 20 nucleosomes [127]. Historically, its role in DNA damage detection has received much attention. More recently, PARP-1 has been linked to the regulation of chromatin structure and transcription (reviewed in [128, 129]). In its enzymatically inactive form, PARP-1 binds chromatin and contributes to the formation of transcriptionally silent chromatin domains [130]. Recent data imply a role in promoting the formation of chromatin structures that are permissive to gene expression [131]. Upon sensing DNA damage, PARP-1 catalyzes the cleavage of its substrate NAD^+ into nicotinamide and ADP-ribose and polymerizes long ADP-ribose chains onto core histones, linker histone H1, and many other nuclear proteins (hetero-modification) as well as onto itself (auto-modification), with itself as the vastly preferred substrate [132]. Mutational studies have revealed several automodification sites in PARP-1 (**Figure 2.1**) [97, 133]. Due to its well-described role in DNA damage repair, PARP-1 is an attractive drug target to augment cancer therapy [134, 135]. However, little quantitative information is available on the many interactions of unmodified and modified PARP-1. For example, it is not known how strongly PARP-1 interacts with nucleosomes compared to nucleosome-free DNA, and whether PARP-1 can recognize DNA damage in the context of chromatin. This limits our understanding of PARP-1 function in chromatin structure maintenance and DNA repair.

PARP-1 contains three N-terminal zinc finger domains, a ‘BRCA1 C-terminus’ (BRCT) domain that is linked to the WGR (‘Tryptophan-Glycine-Arginine-rich’) domain and catalytic domain (CAT) through a flexible linker (**Figure 2.1A**). Structural information on all individual domains is available.

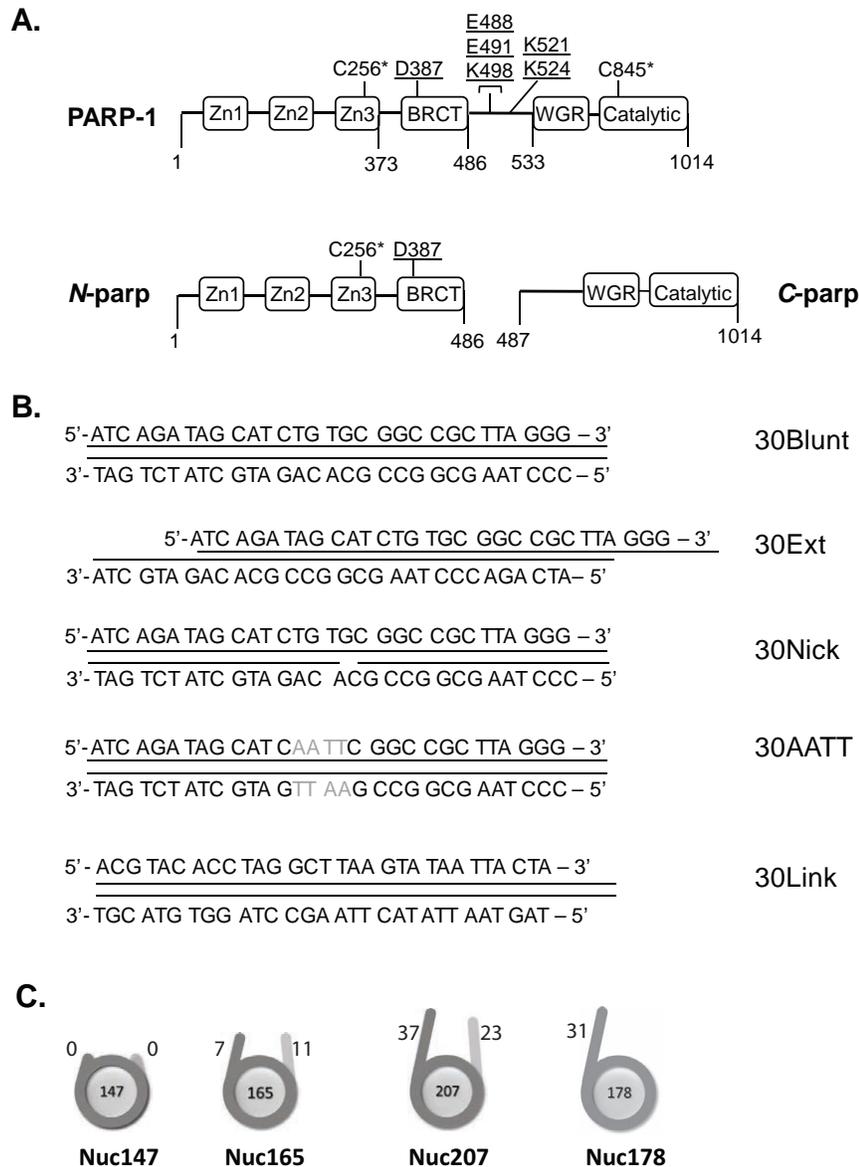


Figure 2.1: PARP-1 constructs and substrates assayed in the present study. **A.** Full length PARP-1 contains all six domains; *N-parp* encompasses Zinc fingers Zn1-Zn3 and the BRCT domain (1-486); *C-parp* spans residues 487-1014 and includes the WGR and catalytic domain. Surface exposed native cysteine residues (positions 256 and 845; indicated by asterisks) were labeled with Alexa488. Underlined residues denote auto-PARYlation sites [97]. **B.** DNA models used for PARP-1 binding and activity assays. 30Blunt, 30Ext, and 30Nick are identical in sequence; 30AATT replaces four central base pairs with AATT. 30Link is identical in sequence to the linker in Nuc178. All DNA models were labeled at the 5' end with Cy5 or Atto647N. **C.** Nucleosome substrates were labeled with Atto647N at H4 E63C on the histone octamer [138]. All nucleosomal DNA is based on the '601' positioning sequence [139]. The length of the linker DNA in each particle is indicated in base pairs.

Zinc fingers 1 and 2 (Zn1 and Zn2) bind DNA with high affinity in a sequence-independent and structure-dependent manner [35, 136], with the strongest interaction observed for Zn2. Zn3 does not bind DNA on its own but is essential for DNA dependent stimulation of PARP-1 activity [137]. It has been proposed that DNA binding by the zinc fingers triggers a conformational change in the full length protein, which then activates the catalytic domain [35]. The impressive structure of a near-full length PARP-1 – DNA complex [49] provides a detailed view of the domain arrangements upon DNA damage, and explains the propensity of PARP-1 for PARylating itself rather than target protein substrates. The crystallized PARP-1 construct, which only lacks Zn2 and the BRCT domain, binds DNA as a monomer, consistent with earlier studies [35, 136] and displays extensive contacts between the DNA damage interface and the catalytic domain. Importantly, the interaction with a single DNA fragment is afforded by residues from Zn1, Zn3, and the WGR domain. This latter domain had previously not been implicated in DNA binding; and it was generally believed that amino acids 1-486 are solely responsible for the interaction with DNA [136, 140]. In addition to recognizing DNA damage, PARP-1 also binds chromatin, and protects an additional ~ 10-20 base pairs of nucleosomal DNA near the entry-exit sites, reminiscent of the pattern observed for H1-nucleosome interaction [123]. Only a moderate contribution of the C-terminal domain of PARP-1 to the interaction with DNA or chromatin was reported [137, 141]. However, this domain is essential for chromatin compaction, independent of its catalytic activity [141].

There are reports that PARP-1 activity is stimulated not only by free DNA but also by chromatin and isolated histones [110, 123, 141]. Consistent with the qualitative observation that PARP-1 also binds mixtures of histones *in vitro*, even in the absence of DNA, PARP-1 is reportedly activated by the N-terminal tail of histone H4 [110]. However, readout of the binding affinities as well as of catalytic activity was indirect. Additionally, no systematic, quantitative comparisons of the degree of PARP-1 activation by the various allosteric activators have been made.

To fill these significant gaps in our understanding of PARP-1 function, we measured the interactions of highly pure full length PARP-1, its N-terminal half (amino acids 1-486, *N-parp*), and its catalytic domain (amino acids 487-1014, *C-parp*) to defined DNA fragments, as well as to nucleosome

substrates with various extensions of linker DNA (**Figure 2.1**). We also quantified the ability of the various binding substrates to stimulate the enzymatic activity of PARP-1. Our data suggest fundamental differences in the mode of interaction between chromatin and free DNA, consistent with the two roles of PARP-1 as a chromatin architectural protein and a sensor of DNA damage. Furthermore, our data demonstrates that PARP-1 is capable of recognizing DNA double strand breaks in the context of a nucleosome.

Experimental Procedures

Expression, purification and fluorescent labeling of PARP-1 and *N*-parp

N-parp was expressed, purified and labeled as described [142]. Full length human PARP-1 V762A was expressed in *Sf9* insect cells [93]. Cell pellets were thawed from -80°C and sonicated (3 x 5 s, output 6.5, duty cycle 65% on a Branson sonifier 450) on ice in lysis buffer (300 mM NaCl, 25 mM Tris/HCl, pH 8, 1 mM β -mercaptoethanol and 1 mM PMSF). Cell lysates were then cleared by centrifugation (14,000 rpm for 30 minutes at 4°C) and the pellet was discarded. DNA was removed by addition of 1.0 mg/ml salmon sperm protamine sulphate (Sigma-Aldrich) followed by centrifugation (14,000rpm for 30 minutes at 4°C). The supernatant was precipitated by a two-step ammonium sulphate treatment at 4°C while stirring overnight. In the first step, the supernatant was incubated with 30% ammonium sulphate (164 g/ 1000 ml), and centrifuged as above. In the second step the supernatant from 30% ammonium sulphate was brought up to 70% ammonium sulphate saturation (249 g/ 1000 ml). The precipitate was resuspended in Heparin chromatography buffer A (100 mM NaCl, 25 mM Tris/HCl, pH 8 and 1.0 mM β -mercaptoethanol), loaded onto a HiTrap Heparin column (GE Healthcare) and eluted with a linear gradient (0-100% B) of Heparin chromatography buffer B (1.5 M NaCl, 25 mM Tris/HCl, pH 8 and 1.0 mM β -mercaptoethanol). Further purification included size-exclusion chromatography and cation

exchange using a HiTrap SP column (GE Healthcare). This homogeneous preparation of PARP-1 tested negatively for automodification by western-blot (data not shown).

Purified full length PARP1 was fluorescently labeled at its native surface exposed (Cys256 and Cys842) cysteine residues. 10 mM Alexa-488 fluorophore (Invitrogen) in DMSO was added to PARP-1 in 300 mM NaCl, 25 mM Tris pH 7.5 in equimolar amounts three times over three hours and allowed to mix overnight at 4°C; excess fluorophore was removed using a Hitrap Heparin HP column as described above. Labeled PARP-1 and *N*-parp run on a 4-12% gradient SDS-PAGE (Criterion XT) appeared as homogeneous bands (Figure 2A, B). A typical labeling efficiency of 10-25% was routinely obtained.

Histone labeling

Histone mutants H4 E63C and H2B T112C were labeled with Atto647N and refolded as described [138] (**Figure 2.2A, B**). A typical labeling efficiency of 10-25% was routinely obtained.

DNA oligomers

30 bp blunt-ended, nicked, and overhang DNA, all containing the template sequence 5'- ATC AGA TAG CAT CTG TGC GGC CGC TTA GGG -3', either with or without a 5'-Cy5 or Atto647N fluorophore were ordered from Integrated DNA Technologies (IDT) (**Figure 2.1B**). Annealing was carried out by mixing equimolar amounts of template and reverse strand, heating at 95° C for 2 minutes, followed by slow cooling to room temperature.

All DNA used for nucleosome assembly contained the 601 positioning sequence with variable linker arms (**Figure 2.1C**) and were expressed and purified as described [143]. The 147 bp DNA represents the minimal nucleosomal DNA; Nuc165 had linker arms of 7 and 11 bp and Nuc207 exhibits linker lengths of 23 and 37 bp, respectively. We also generated an asymmetric linker arm by digesting the 207 bp DNA with BsiEI followed by Mung bean nuclease digestion producing the 178 bp DNA (**Figure 2.1C**).

Chromatin Assembly and Characterization

Labeled nucleosomes were assembled on DNA of varying lengths as described [143] using Atto647N labeled histone octamer (**Figure 3A, B**). The nucleosome preparations typically had <1% free DNA present.

HI-FI FRET Assay

We used the previously developed HI-FI FRET [142] for measuring affinities and stoichiometries for PARP-1 and *N*-parp labeled with donor dye Alexa 488 and titrated in substrates labeled with acceptor dye Atto647N. Buffer used for setting up the binding reactions included 25 mM Tris pH 7.5, 200 mM NaCl, 0.01 % (v/v) each of NP40 and CHAPS. FRET calculations and corrections were performed as described [142]. The data were plotted in Graphpad Prism and fitted using one site binding + background or one site-specific binding with Hill slope. The data were represented by plotting titrated species labeled with the acceptor on the X-axis and normalized FRET corrected values on the Y-axis. The Hill coefficient was held constant at 1 unless otherwise mentioned.

Electrophoretic Mobility Shift Assay (EMSA)

Labeled Nuc165 (1 μ M) was titrated with increasing molar ratios of PARP-1 or *N*-parp labeled with Alexa488 in the binding buffer described above and incubated for 30 minutes at room temperature. Samples were subsequently run on a 22 cm x 20 cm native TBE gel and run in 0.5X TBE, at 4°C at 300V, 10W for 120 minutes. The gel was scanned on a Typhoon Imager at wavelengths appropriate for measuring acceptor (633 excitation; 670 emission), donor (488 excitation; 520 emission) and FRET (488 excitation; 670 emission). Gels were then stained with Ethidium Bromide to visualize the DNA.

Unlabeled nucleosomes (1 μ M) were incubated with increasing amounts of labeled or unlabeled PARP-1 constructs (PARP-1, *N*-parp and *C*-parp) in 25 mM or 50 mM Tris pH 7.5, 150 mM NaCl, 2 mM

Arginine, 0.01% CHAPS and NP40. The DNA/chromatin-PARP-1 samples were incubated at room temperature for 30 minutes, loaded on a pre-run 5% native TBE gel and run at 150 V for 60 minutes at 4°C for 8 cm x 8 cm gels in 0.2X TBE. Gels were stained with Ethidium Bromide followed by Imperial stain for protein.

Size Exclusion Chromatography/Multi-Angle Light Scattering (SEC-MALS)

Nucleosomes (Nuc147, Nuc165 and Nuc207) and their complexes with PARP-1 were assembled in 50 mM Tris pH 7.5, 150 or 300 mM NaCl, 2 mM Arginine and analyzed in SEC-MALS as described [144].

PARP-1 Enzymatic Assay

PARP-1 (constant at 1 μ M) and ‘activators’ (DNA or nucleosomes; 1-2 μ M) were mixed to a final volume of 30 μ l in 50 mM Tris pH 8, 50 mM NaCl (100 mM NaCl for chromatin activators), 10 mM MgCl₂ (or 1 mM MgCl₂ for chromatin activators) and 1 mM DTT and allowed to incubate for 1 hour at 30°C. 30 μ l of the various NAD⁺ stocks (0-400 μ M) were added to the above tubes. Reactions were quenched after 30 seconds with either Laemmli buffer or ice cold 20% trichloroacetic acid (TCA). Reactions quenched with Laemmli buffer were analyzed on 8% SDS-PAGE and by western blot. 1-5% of reactions quenched with 20% TCA were loaded onto a Zeta-probe membrane (BioRad) using a BioRad[®] Dot Blot apparatus [93]. A PAR standard curve was also included in each blot to correlate the amount of PAR generated by automodification directly to a known amount of standard PAR. After loading the sample, the wells were washed once with 10% TCA followed by washing with 70% ethanol. The membrane was then dried on a gel dryer at 80° C for 1 hour, and blocked with 5% milk in 1x TBS overnight. The blot was incubated in primary anti-PAR antibody (Abcam) for 1 hour followed by five washes with 1x TBS + 0.01% v/v Tween 20. Secondary goat anti-mouse antibodies conjugated with Atto647 (Sigma) were incubated for 1 hour followed by five washes of 1x TBS containing 0.01% Tween

20. The blots were imaged on a Typhoon Imager using the Acceptor $\lambda = 633$ nm Excitation; 670 nm Emission. (GE Healthcare) and quantified using the array analysis program of ImageQuant. Michaelis-Menten parameters were derived using the GraphPad Prism v5[®] non-linear regression.

Results

PARP-1 exhibits a slight preference for flexible DNA

We have previously shown by agarose-gel mobility shift assays that a fragment of PARP-1 encompassing the three Zinc fingers and the BRCT domain (*N*-parp; **Figure 2.1A**) binds tightly to various DNA damage models [136]. We wanted to investigate how full length PARP-1 compares to *N*-parp, using a more rigorous solution-state assay that we recently developed in our lab [142, 145]. PARP-1 and *N*-parp were purified to homogeneity and labeled with fluorophores as described [142] (Figure 2A, B). Electrophoretic mobility shifts were observed when a 30 base pair DNA fragment (30Blunt DNA) was titrated with either full length PARP-1 or *N*-parp, qualitatively confirming that both fluorescently labeled proteins form defined complexes with DNA (data not shown). Quantitative information on the interactions was obtained by monitoring binding reactions through Fluorescence Resonance Energy Transfer (FRET) in a plate assay (HI-FI FRET) [142]; representative data is shown in **Figures 2.2C, D**. **Table 2.1** summarizes the affinities of the two PARP-1 constructs for the free DNA models listed in Figure 1B. Both full length and the N-terminal half of PARP-1 exhibit a slight preference for DNA containing an internal nick or an AATT insert. These features are thought to induce a curved or bent conformation into double stranded DNA [146].

The dissociation constants of *N*-parp – DNA complexes, as determined by HI-FI FRET, compare well to the previously reported affinities for the various DNA damage models 30Blunt, 30Ext and 30Nick DNA [136]. The overall 3-5 fold tighter affinities of *N*-parp in the current study are likely due to differences in binding conditions (200 mM NaCl here versus 300 mM NaCl in previous studies). This is

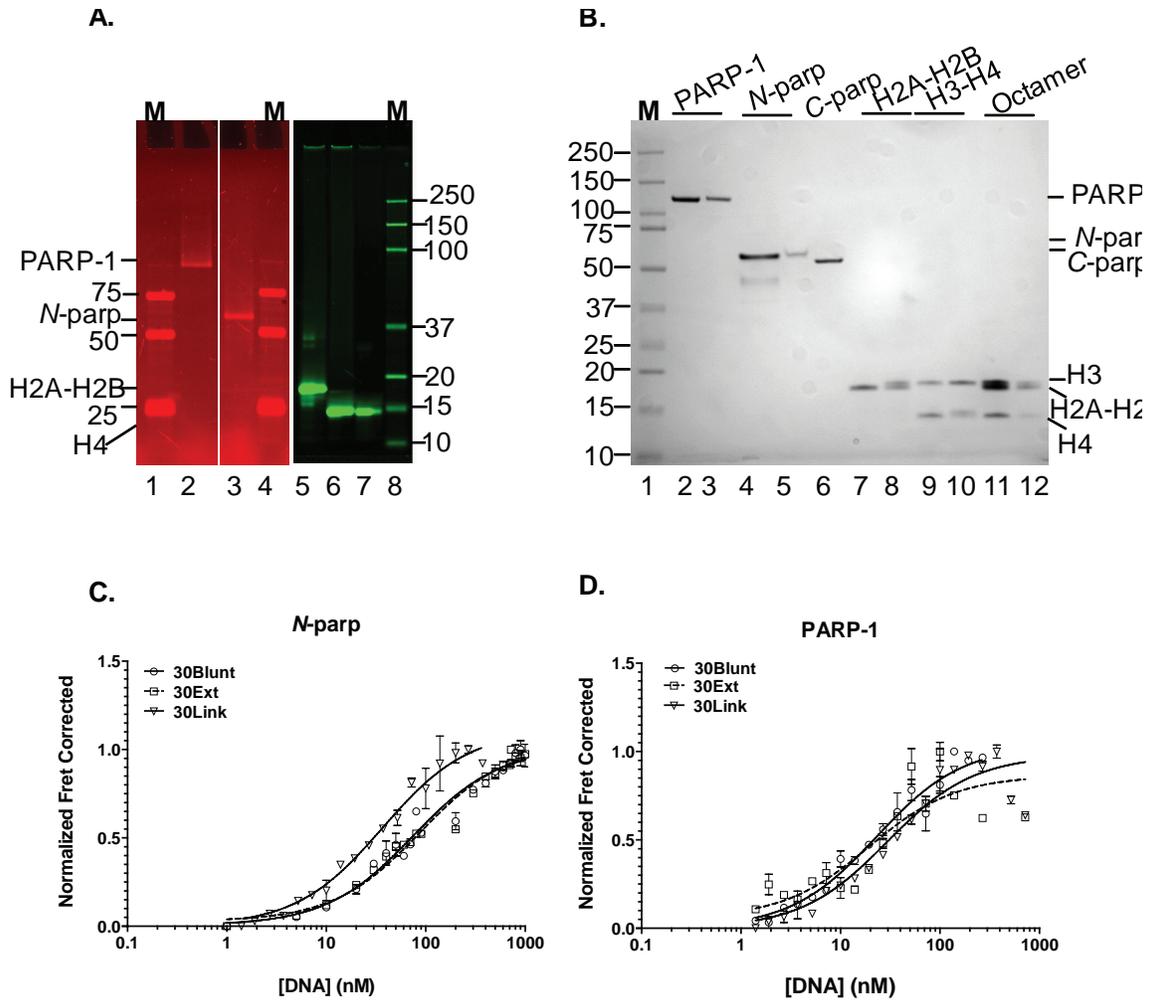


Figure 2.2. The catalytic domain of PARP-1 contributes moderately to the interaction with DNA. A. Fluorescently labeled PARP-1 constructs and histones. All samples were run on a Criterion XT 4-12% gradient SDS-PAGE gel that was scanned on a Typhoon Imager at wavelengths appropriate for measuring donor (488 excitation; 520 emission) for the left panels in A depicting PARP-1 constructs, and acceptor (633 excitation; 670 emission) for the right panel with labeled histones. Lanes 1, 4, 8: molecular weight marker; lane 2: Alexa488 labeled PARP-1 (at Cys256 and Cys845); lane 3: Alexa488 labeled *N*-parp (at Cys256); lane 5: H2A-H2B dimer (with Atto647N-labeled at H2B T112C); lane 6: (H3-H4)₂ tetramer (labeled with Atto647N at H4E63C); lane 7: histone octamer (labeled with Atto647N at H4E36C). **B.** The same gel was visualized with Imperial stain. Lane 1: protein size marker; lane 2: unlabeled PARP-1; lane 3: labeled PARP-1; lane 4: unlabeled *N*-parp; lane 5: labeled *N*-parp; lane 6: *C*-parp; lane 7: unlabeled H2A-H2B; lane 8: labeled H2A-H2B; lane 9: unlabeled H3-H4; lane 10: labeled H3-H4; lane 11: unlabeled histone octamer, Lane 12: labeled histone octamer. **C.** *N*-parp (labeled with Alexa488 at Cys256) binding to selected free DNA models shown in Figure 1B, as measured by HI-FI FRET. **D.**

Table 2.1. Relative affinities of *N*-parp, and PARP-1 for various free DNA models and nucleosomes. Data were obtained using the HI-FI FRET assay [142]. Buffer for all binding reactions contained 25 mM Tris pH 7.5, 200 mM NaCl, 0.01% (v/v) each NP40 and CHAPS (with the exception of the values indicated by *). Standard deviations are reported for 2-5 independent experiments, except for the value indicated by **.

* 250 mM NaCl instead of 200 mM NaCl. At 200 mM NaCl the affinity of PARP-1 for 30Nick was in the low nanomolar range (data not shown).

**Errors are derived from one data set only

| Binding substrate | <i>N</i>-parp | | PARP-1 | |
|--------------------------|------------------------------------|-------------------------|------------------------------------|-------------------------|
| | K_d^{app} (nM) | R^2 | K_d^{app} (nM) | R^2 |
| 30Blunt | 62.2 ± 10.2 | 0.97 | 31.7 ± 6.9 | 0.95 |
| 30Ext | 111.5 ± 30.5 | 0.98 | 66.0 ± 11.0 | 0.89 |
| 30Nick | 27.8 ± 5.6 | 0.95 | 23.4 ± 4.8* | 0.98 |
| 30AATT | 25.7 ± 0.9 | 0.92 | 8.5 ± 2.1 | 0.87 |
| 30Link | 33.1 ± 1.5 | 0.98 | 24.0 ± 2.0 | 0.96 |
| Nuc147 | > 500 | 0.99 | >500 | 0.98 |
| Nuc165 | 57.8 ± 6.1 | 0.88 | 2.2 ± 1.5 | 0.86 |
| Nuc207 | 48.8 ± 21.2 | 0.97 | 1.0 ± 0.2 | 0.90 |
| Nuc178 | 238.0 ± 26.5 | 0.92 | 84.6 ± 7.7** | 0.98 |

in keeping with the previously observed strong dependence of PARP-1-DNA interactions on ionic strength [142]. Compared to *N*-parp, full length PARP-1 exhibits 1.4-3 fold tighter affinity for all free DNA models (**Table 2.1**). This indicates that the C-terminal half of PARP-1 contributes moderately to the binding event, consistent with structural data demonstrating interactions between the WGR domain (not contained in *N*-parp) and DNA [49]. The C-terminal half of PARP-1 on its own is unable to interact measurably with DNA (data not shown).

A single PARP-1 molecule interacts strongly with a nucleosome containing symmetric linker DNA

We next wanted to test *N*-parp and PARP-1 affinities for defined mono-nucleosomes that vary in length and symmetry of their linker DNA (**Figure 1C**). Nuc147 is a mono-nucleosome that completely lacks DNA linker arms, while Nuc165 and Nuc207 contain two linker arms each. Nuc178 was designed to have only one exposed linker arm. The sequence of this 30 base pair extension is identical to that of 30Link (**Figure 2.1**). According to our analysis by native PAGE, all nucleosomes are uniquely positioned, and the percentage of free DNA in each of these nucleosome preparations was below 1% (**Figure 2.3A, B**). The addition of fluorophore to histones does not change the electrophoretic mobility of the reconstituted nucleosomes, indicating that they are structurally intact. The interaction of *N*-parp and PARP-1 with Nuc165 was first tested by EMSA. When fluorescently labeled Nuc165 was titrated with either labeled PARP-1 or *N*-parp, distinct bands that exhibit both acceptor and donor fluorescence were observed (**Figure 2.3C**). These bands also displayed FRET (pink bands in bottom overlay panel), providing further proof of defined complex formation.

We next quantified the interaction of *N*-parp and PARP-1 with the various nucleosome substrates in solution, using HI-FI FRET (**Figure 2.4A**). Nuc165 as well as Nuc207 bind *N*-parp with 50-60 nM affinity, while no plateau was achieved with Nuc147 (**Figure 2.4B**), characteristic of very weak interaction. Because regions outside of *N*-parp are known not to interact with DNA on their own, we were surprised to see that full length PARP-1 binds nucleosomes with two DNA linker ends 25-50-fold tighter than *N*-parp (**Figure 2.4C**). In light of the moderate difference of the binding affinity of the two PARP-1

constructs for free DNA, this suggests a substantial contribution of the catalytic domain to the interaction with nucleosomes containing two DNA linker arms. This is despite the inability of the C-terminal domain of PARP-1 (*C-parp*) to bind mono-nucleosomes when tested by EMSA (**Figure 2.4D**). Like *N-parp*, full length PARP-1 binds Nuc147 only weakly (**Figure 2.4C**). Importantly, interaction of full length PARP-1 with Nuc207 and Nuc165 is significantly tighter than to any of the free DNA substrates.

To further test whether both DNA linker arms are required for a stable PARP-1 interaction, we generated a nucleosome with a single asymmetric ~30 bp extension of DNA linker (Nuc178; **Figure 2.1C**). Both PARP-1 constructs bind this nucleosome substrate with significantly reduced affinity compared to Nuc165 or Nuc207 (**Table 2.1**). The data suggests that both linker arms are required for optimal PARP-1 binding. Binding of both PARP-1 constructs to Nuc178 is also 3-7-fold weaker than to the corresponding ‘free’ 30mer with identical sequence (30Link), indicating steric hindrance of PARP-1 binding to Nuc178. Finally, full length PARP-1 binds these nucleosomes only 3-fold tighter than does *N-parp*, in contrast with the 25-50- fold increase of affinity for nucleosomes with two linker ends. This suggests that the catalytic domain contributes to positioning PARP-1 in a way that allows engagement of both DNA linker arms, and thus does not contribute as much to the interaction with a nucleosome with only one linker arm.

We next wanted to determine the stoichiometry of the various nucleosome-PARP-1 complexes. Nuc207, Nuc165, or Nuc147 were mixed with varying amounts of PARP-1 and analyzed by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS; **Figure 2.5 and Table 2.2**). For the complexes between Nuc207 or Nuc165 and PARP-1, the observed molecular weights matched the calculated value for a 1:1 complex even when excess PARP-1 was added. In this case, a second peak for free PARP-1 was observed. A stoichiometry of 1:1 was also measured for *N-parp*-nucleosome complexes (data not shown). Consistent with the low binding affinity, Nuc147 and PARP-1 eluted as two separate peaks (**Table 2.1**), despite the residual interactions observed by native PAGE (**Figure 2.4D**).

Together, our quantitative analysis of PARP-1 nucleosome binding and stoichiometry reveals a strong contribution of the WGR-CAT domain to the interaction with nucleosomes with two linker arms

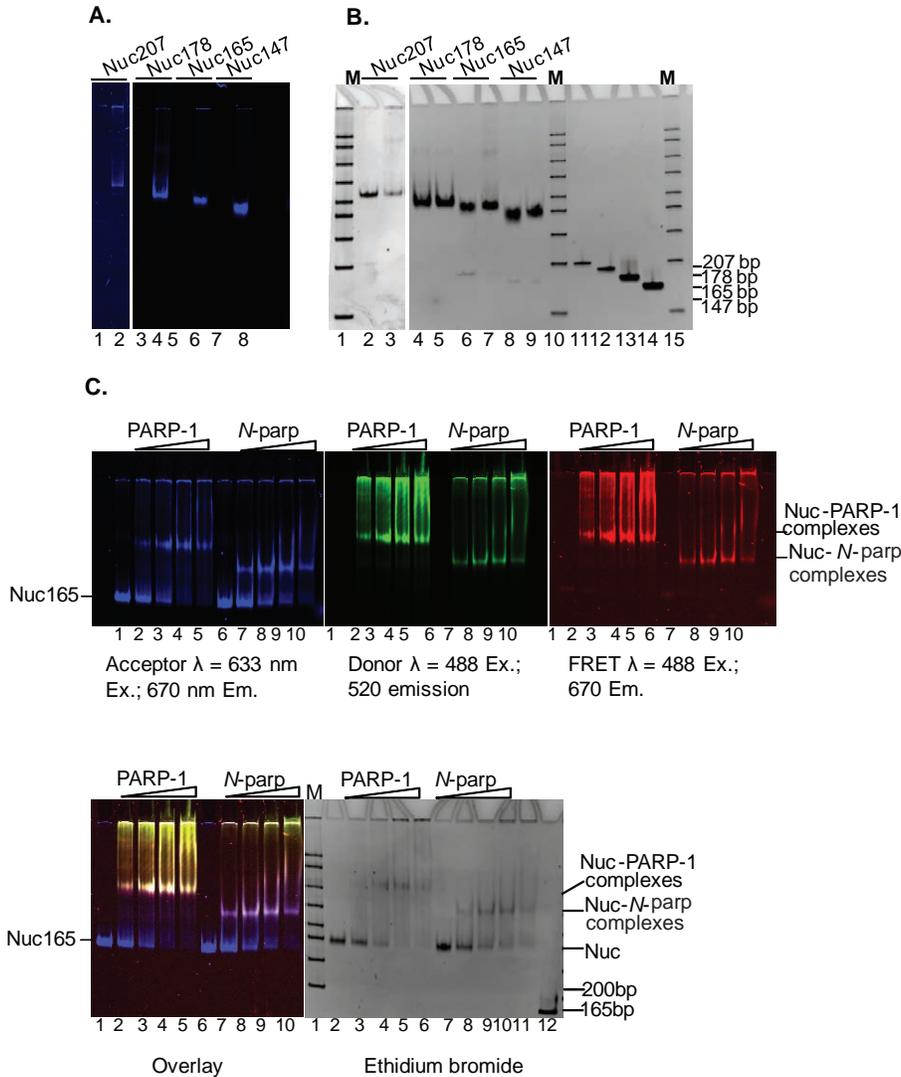


Figure 2.3. PARP-1 interacts with nucleosomes. A. Fluorescently labeled nucleosome substrates: DNA fragments 207, 178, 165 and 147 base pairs in length, all containing the 601 positioning sequence, were assembled into nucleosomes with histone octamers labeled at H4E63C with Atto647N. Nucleosomes were run on 5% native PAGE and scanned on a Typhoon imager at an emission wavelength of 670 nm. Lanes 2, 4, 6 and 8 are nucleosomes assembled on 207, 178, 165 and 147 bp DNA, respectively. **B.** The same gel, stained with ethidium bromide. Lanes 1 and 2 are labeled and unlabeled Nuc207 respectively, lanes 3 and 4 are labeled and unlabeled Nuc178, lanes 5 and 6 are labeled and unlabeled Nuc165, lanes 7 and 8 are labeled and unlabeled Nuc147. Lanes 9-12 are 207, 178, 165 and 147 bp DNA fragments, respectively. Note the absence of free DNA (<1%) in the nucleosome samples. **C.** Atto647N (acceptor)-labeled nucleosomes (Nuc165) were incubated with increasing amounts of Alexa488-labeled PARP-1 or N-parp, and analyzed by native PAGE. Gels were scanned on a Typhoon Imager at the indicated wavelengths. Overlay: Acceptor, Donor and FRET channels are overlaid. Lanes 1 and 6: Nuc165; lanes 2-5: nucleosomes incubated with increasing molar ratios of PARP-1 (0.5, 1, 1.5 and 2 fold excess). Lanes 7-10: nucleosomes incubated with increasing molar ratios of N-parp (0.5, 1, 1.5 and 2 fold excess). Lane 11 in the ethidium bromide panel: free 165 bp DNA.

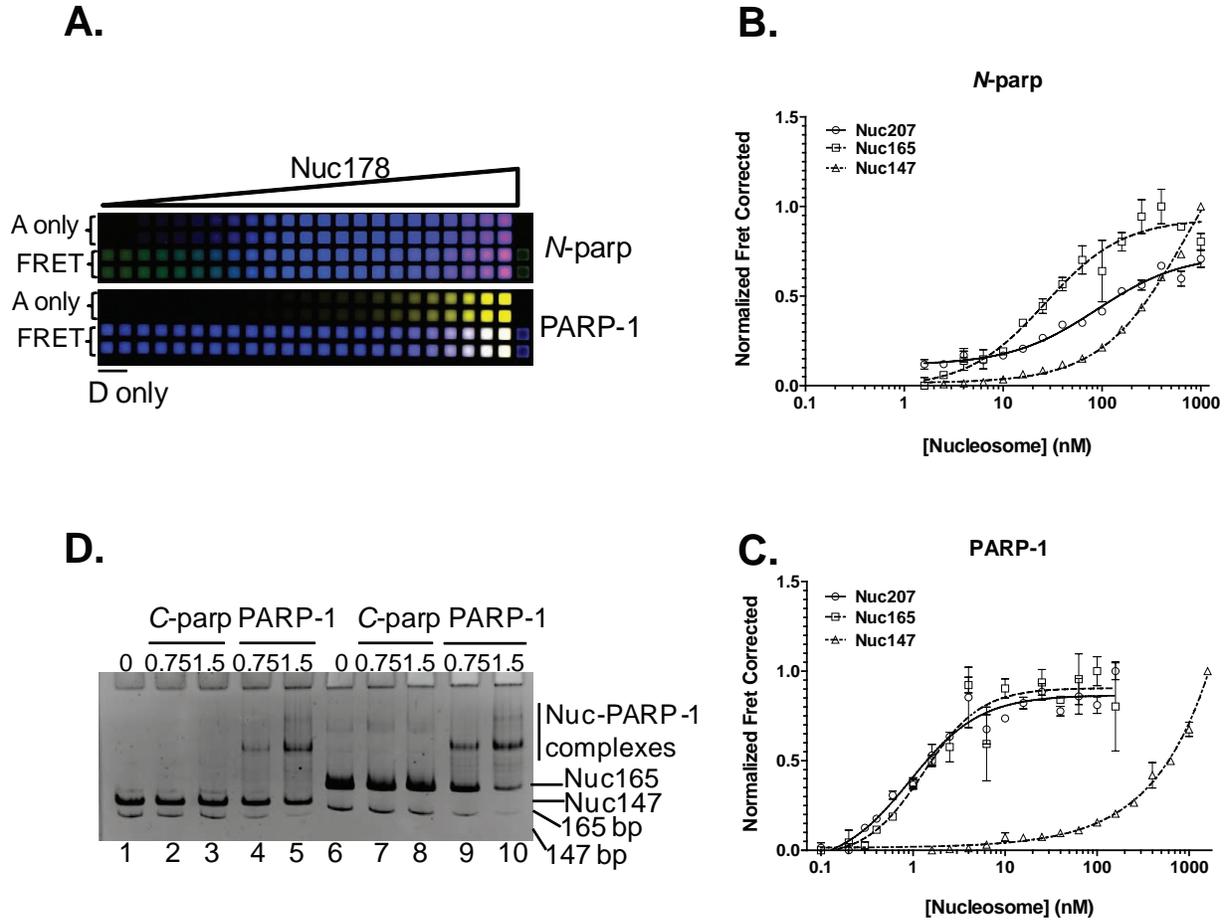


Figure 2.4: Quantification of interactions between PARP-1 and nucleosomes. **A.** Hi-Fi FRET plate assay. A portion of a typical 384-well plate is shown for Nuc178 and *N-parp* (top panel) and full length PARP-1 (bottom panel). Increasing amounts of Nuc178 labeled with Atto647N at H4E63C was titrated with constant amount of either *N-parp* or PARP-1 labeled with Alexa488. The top two rows in each panel represent acceptor only (A only) controls. The first two wells in bottom two rows in each panel are donor only (D only) wells. FRET between the interacting partners is shown in the bottom two rows of each panel (pink/purple). The plate was scanned using a Typhoon imager as described for the gel in Figure 2. Data from experiments were normalized, and the resulting curves were fit as described [142]. Results from this plate are shown in **Table 2.1**. **B.** *N-parp* and **C.** PARP-1 interactions with the various mono-nucleosome substrates. All values from this and similar experiments are summarized in **Table 2.1**. **D.** *C-parp* does not bind nucleosomes. Nucleosomes were incubated with *C-parp* or PARP-1 at increasing molar excess (as indicated) and loaded on a pre-run 5% native TBE gel. Gels were stained with Ethidium bromide. *C-parp* does not interact with Nuc147 (lanes 2 and 3) or Nuc165 (lanes 7 and 8), while PARP-1 causes an upshift in both (lanes 4, 5 for Nuc147, and 9, 10 for Nuc165).

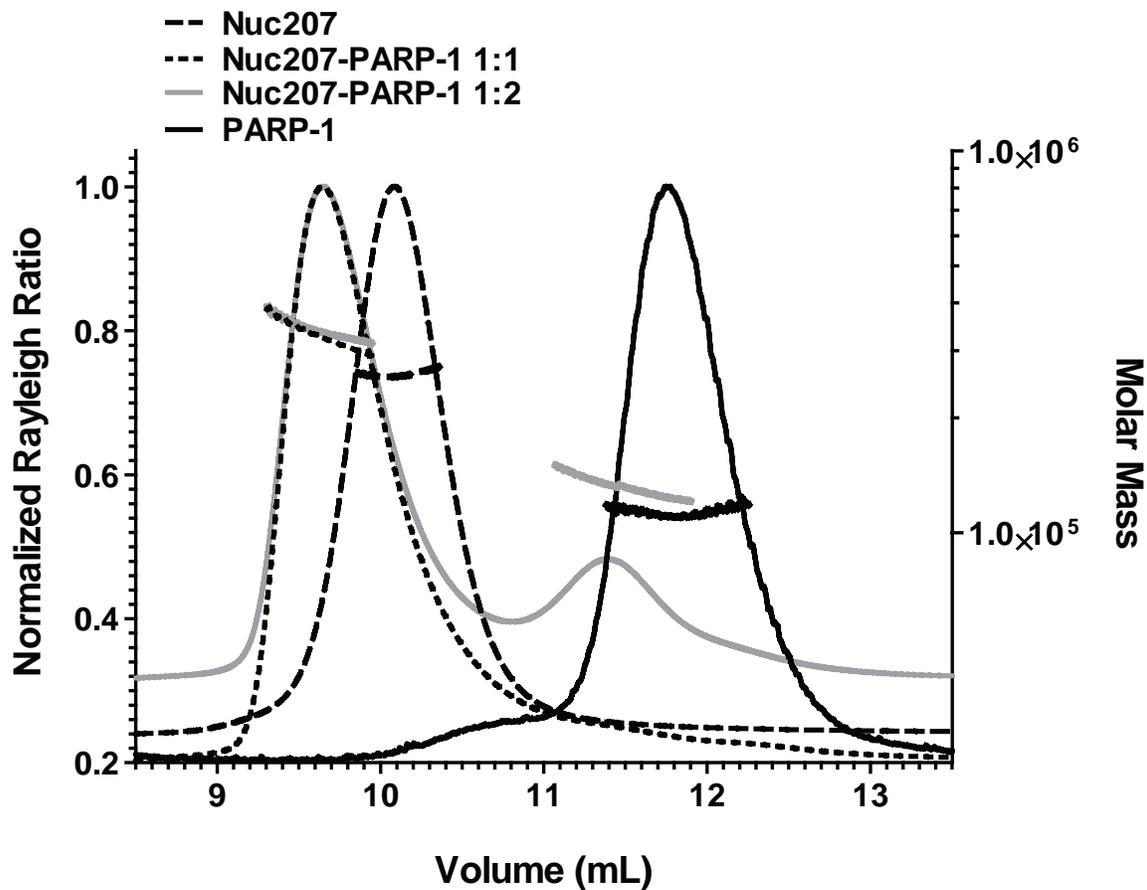


Figure 2.5: One PARP-1 molecule binds per nucleosome. Size exclusion chromatography – multi-angle light scattering (SEC-MALS) profiles for Nuc207 and its complexes with PARP-1. Nuc207 forms a 1:1 complex with PARP-1 even when excess PARP-1 is added to the reaction mix. The molecular weights for the various complexes derived from this and similar SEC-MALS experiments are listed in **Table 2.2**.

Table 2.2. Molecular mass analysis of nucleosomes and their complexes with PARP-1 as determined by SEC-MALS (Figure 5).

| | | Observed MW | Calculated MW |
|---------------|-----------------------|--------------------|----------------------|
| Nuc207 | Nuc207 | 2.45×10^5 | 2.36×10^5 |
| | PARP-1 | 1.1×10^5 | 1.13×10^5 |
| | Nuc207 + PARP-1 (1:1) | 3.5×10^5 | 3.49×10^5 |
| | Nuc207 + PARP-1 (1:2) | 3.5×10^5 | 4.62×10^5 |
| Nuc165 | Nuc165 | 2.17×10^5 | 2.10×10^5 |
| | PARP-1 | 1.01×10^5 | 1.13×10^5 |
| | Nuc165 + PARP-1 (1:1) | 2.6×10^5 | 3.23×10^5 |
| | Nuc165 + PARP-1 (1:2) | 3.08×10^5 | 4.4×10^5 |
| Nuc147 | Nuc147 | 2.02×10^5 | 1.99×10^5 |
| | PARP-1 | 1.01×10^5 | 1.13×10^5 |
| | Nuc147 + PARP-1 (1:1) | 1.95×10^5 | 3.12×10^5 |
| | Nuc147 + PARP-1 (1:2) | 1.82×10^5 | 4.25×10^5 |

(25-50 fold increased affinity), whereas the contribution of these domains to the interaction with free DNA is moderate at best (1.4-3-fold). Similarly, the affinity of full length PARP-1 for Nuc178 is only increased three-fold compared to *N-parp*. Since nucleosomes without linker DNA show no significant PARP-1 binding, we conclude that the contributions of the ‘nucleosome core’ itself are minimal. Thus, high-affinity binding of full length PARP-1 is provided by specific arrangement of two linker DNA arms that are only provided in the context of a nucleosome.

PARP-1 is activated by DNA and nucleosomes

In light of the tight interaction of PARP-1 with a variety of DNA and chromatin substrates, we wanted to know what triggers the catalytic activity of PARP-1, and whether there is a quantitative difference in the degree of activation by the different DNA and chromatin substrates. To address these questions, we measured PARP-1 activity in the presence of various DNA and chromatin activators using the slot blot method [93]. A representative case of PARP-1 activation by 30Blunt DNA is shown in **Figure 2.6**. SDS-PAGE followed by western blot analysis with anti- poly-ADP ribose (PAR) antibody clearly demonstrates an up-shift in the PARP-1 band with increasing NAD^+ concentrations, indicative of the addition of PAR chains to PARP-1 (**Figure 2.6A**). To quantify the amount of PAR generated in each reaction, samples were analyzed by a slot blot and probed with the same anti-PAR antibody used above (**Figure 2.6B**). The data was plotted in GraphPad Prism using Michaelis-Menten curve fitting (Figure 6C). The enzymatic parameters for PARP-1, in the absence and presence of the various activators are summarized in **Table 2.3**. k_{cat} in the absence of DNA reflect the low basal background activity of PARP-1.

PARP-1 is significantly activated over background levels by all linear DNA substrates, as evident by increases in V_{max} ; our values are in good agreement with those obtained using a similar approach [93]. Closed circular plasmid DNA causes residual enzyme turnover, presumably due to the unavoidable contamination with nicked or linear DNA in most plasmid preparations. While PARP-1 binds NAD^+ even

in the absence of DNA (no significant changes in K_M), k_{cat} values range between 0.9-2 per second for all linear DNA fragments, but are near 0 in the absence of DNA (**Table 2.3**). Nucleosomes with either one or two linker ends (Nuc178 and Nuc207) activate PARP-1 to a similar degree, despite the difference in binding affinity and presumably binding mode. Nuc147, which lacks free linker ends, has only reduced ability to stimulate PARP-1. Chromatin with at least one free DNA end activates PARP-1 to a higher degree than a linear DNA fragment with the same sequence (compare 30Link and Nuc178; **Table 2.3**). While Nuc147 is rather inefficient at activating PARP-1 at 100 mM NaCl, it becomes a better activator at 50 mM NaCl, consistent with the idea that lower ionic strength promotes spontaneous “breathing” of the DNA ends [147, 148]. The degree of activation resembles that achieved by short free DNA segments under the same conditions, where Nuc178 is still superior to either substrate. No PARP-1 activation was observed in the presence of any of the histone sub-complexes in the absence of DNA (data not shown). Thus, under our conditions, there is no direct correlation between activation and binding affinity; however, the presence of a nucleosome in addition to a DNA double strand break appears to contribute to PARP-1 activation.

PARP-1 is a highly abundant nuclear protein with a multitude of biological functions (reviewed in [129]). PARP-1 contributes to the compaction of chromatin through direct interactions with nucleosomes, but also binds various forms of damaged DNA. While its interaction with free DNA has been reasonably well-studied (e.g. [35, 49, 136], much less is known about the interaction of PARP-1 with chromatin. To fill this significant gap, we measured the affinity of PARP-1 for defined DNA and chromatin substrates, and quantified the degree of stimulation of its enzymatic activity by the various for high-affinity binding to nucleosomes; and iii) a requirement for at least one free DNA end on the nucleosome for enzymatic activation.

Discussion

Our analysis of the interaction of PARP-1 with short DNA fragments reveals that PARP-1 prefers

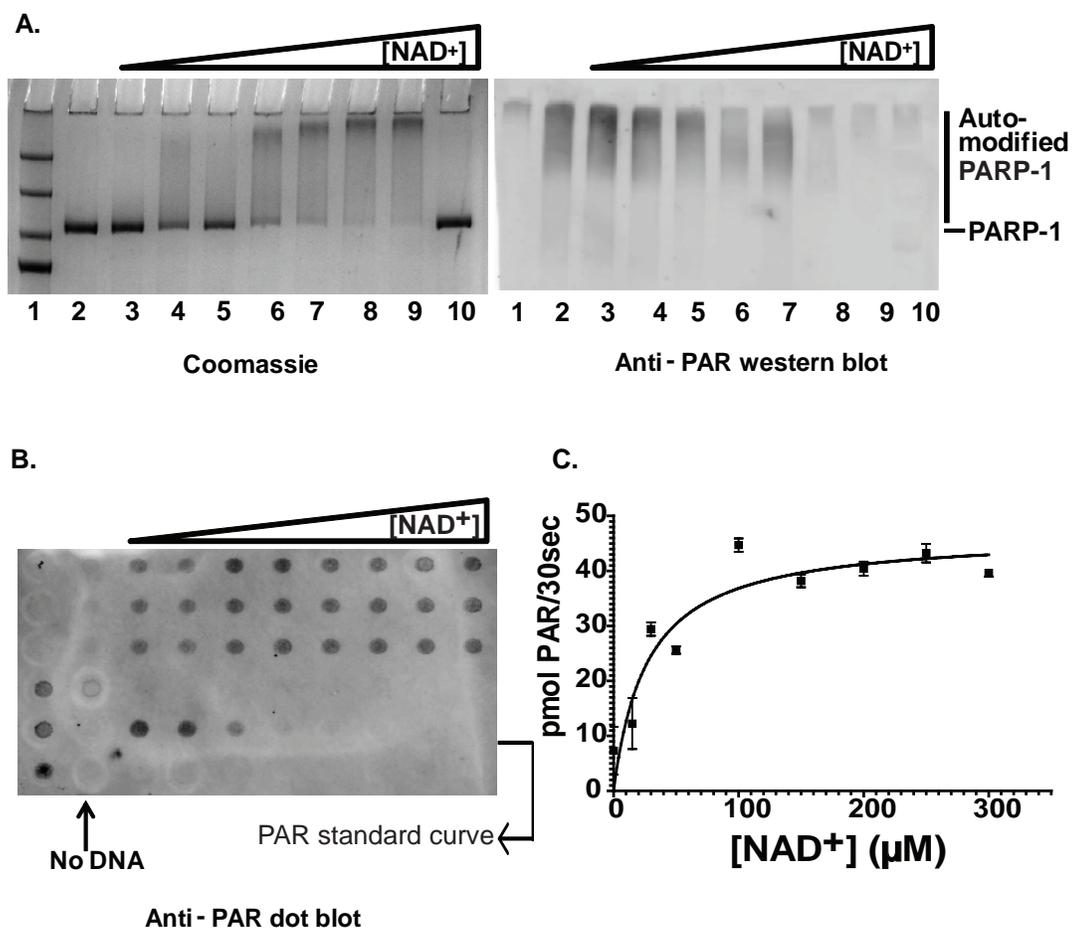


Figure 2.6: PARP-1 is activated by DNA. **A.** SDS-PAGE showing a shift in PARP-1 mobility as it undergoes auto-PARylation in the presence of 1 μM 30Blunt DNA and increasing concentrations (0, 10, 20, 50, 100, 200 and 400 μM in lanes 3-9 respectively) of NAD^+ . The right-hand panel is a western blot of an identical gel probed with anti-PAR antibodies. Lane 1: protein size marker. Lane 2: no NAD^+ ; Lanes 3-9 have increasing amounts of NAD^+ ; lane 10: no DNA in the presence of 400 μM NAD^+ . **B.** A slot blot of the above reaction was probed with anti-PAR antibody and with secondary antibodies conjugated with Atto647, and visualized on a Typhoon scanner at 633 nm Excitation; 670 nm emission wavelength. **C.** The data were quantified using ImageQuant TL and analyzed in a Michaelis-Menten plot. A complete list of all parameters for this and other activators are listed in **Table 2.3**.

Table 2.3. Enzymatic parameters for PARP-1 upon activation by DNA or nucleosomes. *K_d values are taken from **Table 2.1**.

Reaction conditions : DNA[50]=50 mM Tris 8.0, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT

DNA[100]=50 mM Tris 8.0, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT

Chromatin[50]=50 mM Tris 8.0, 50 mM NaCl, 1 mM MgCl₂, 1 mM DTT

Chromatin[100]=50 mM Tris 8.0, 100 mM NaCl, 1 mM MgCl₂, 1 mM DTT

| Allosteric Activators | K_d^{app} (nM)* | V_{max} (pmol/min μg) | K_M (μM NAD⁺) | k_{cat} (s⁻¹) | k_{cat}/K_M (s⁻¹M⁻¹) |
|------------------------------|--|--------------------------------------|---|---|---|
| No DNA | | | | | |
| | | 9.40 | 19.4 ± 30.8 | 0.02 | 0.09 |
| DNA [50] | | | | | |
| pGEM-3Z | | 188.0 | 62.0 ± 17.7 | 0.35 | 0.56 x 10 ⁴ |
| 30Blunt | 32 | 541.1 | 27.1 ± 6.03 | 1.02 | 3.95 x 10 ⁴ |
| 30Ext | 66 | 541.4 | 62.9 ± 9.9 | 1.01 | 4.31 x 10 ⁴ |
| 30Nick | 23 | 311.0 | 33.0 ± 7.6 | 0.88 | 1.89 x 10 ⁴ |
| 30AATT | 8.5 | 568.2 | 25.7 ± 6.4 | 1.06 | 6.55 x 10 ⁴ |
| 30Link | 24 | 1070.6 | 33.9 ± 18.9 | 1.99 | 6.03 x 10 ⁴ |
| Chromatin [50] | | | | | |
| Nuc147 | | 642.4 | 23.1 ± 7.52 | 1.2 | 5.20 x 10 ⁴ |
| Nuc178 | | 1408.2 | 16.4 ± 4 | 2.65 | 16.15 x 10 ⁴ |
| Chromatin [100] | | | | | |
| Nuc147 | >500 | 82.3 | 37.4 ± 21.0 | 0.15 | 0.42 x 10 ⁴ |
| Nuc207 | 1.0 | 786.4 | 23.6 ± 8.0 | 1.47 | 6.22 x 10 ⁴ |
| Nuc178 | 85.0 | 709.8 | 7.6 ± 10.5 | 1.33 | 1.76 x 10 ⁴ |
| DNA [100] | | | | | |
| 30Blunt | | 148.1 | 20.6 ± 9.3 | 0.28 | 1.40 x 10 ⁴ |

ligands. Together, our data demonstrate, i) a significant contribution of the WGR-CAT domains to the interaction of PARP-1 with nucleosomes, but not to free DNA; ii) a requirement for a pair of linker DNA DNA substrates with a propensity to bend. EM studies have shown that PARP-1 induces a bend into nicked or gapped DNA [61]. The recognition of the weakened base stacking and the increased flexibility at DNA lesion sites has been proposed as a first step in DNA damage recognition by many repair proteins [149]. Consistent with previous qualitative reports [123] and with the recent crystal structure of near full length PARP-1 in complex with a short DNA fragment [49], we find a modest contribution of the C-terminal half of PARP-1 (presumably due to the interactions made by the WGR domain) to the interaction with each of the short DNA fragments tested.

Full length PARP-1 binds very tightly to mono-nucleosomes that contain at least 10 base pairs of linker DNA extending on either side. This is consistent with the result that 160 base pairs of nucleosomal DNA are protected from MNase digestion in the presence of PARP-1 [123], but contradicts indirect evidence that linker DNA does not contribute to the interaction [110]. The strong contribution of the C-terminal half of PARP-1 to the interaction with nucleosomes is striking because this region on its own does not measurably bind nucleosomes, and because full length PARP-1 interacts only very marginally with nucleosomes lacking DNA linkers (Nuc147). Since histones are also subject to a low degree of PARylation, this interaction might also entail substrate recognition by the catalytic domain. However, the interaction of PARP-1 with nucleosomal histones or histone tails is not sufficient for robust binding in the absence of linker DNA.

PARP-1 likely engages both DNA linker ends, since the deletion of one of the two linker arms in a nucleosome (leaving one 30 base pair DNA linker, Nuc178) resulted in a significant reduction in binding, and in a much reduced contribution of WGR-CAT to the interaction. The model for the structure of full length PARP-1 in complex with a single DNA fragment proposed by Langelier and colleagues [49], suggests that Zn2 might be the domain responsible for interacting with the second DNA linker end (**Figure 2.7**). This domain binds linear DNA on its own, and with tighter affinity than Zn1, but is not needed for the enzymatic activation of PARP-1 [137]. The 25-50-fold increase in binding affinity for

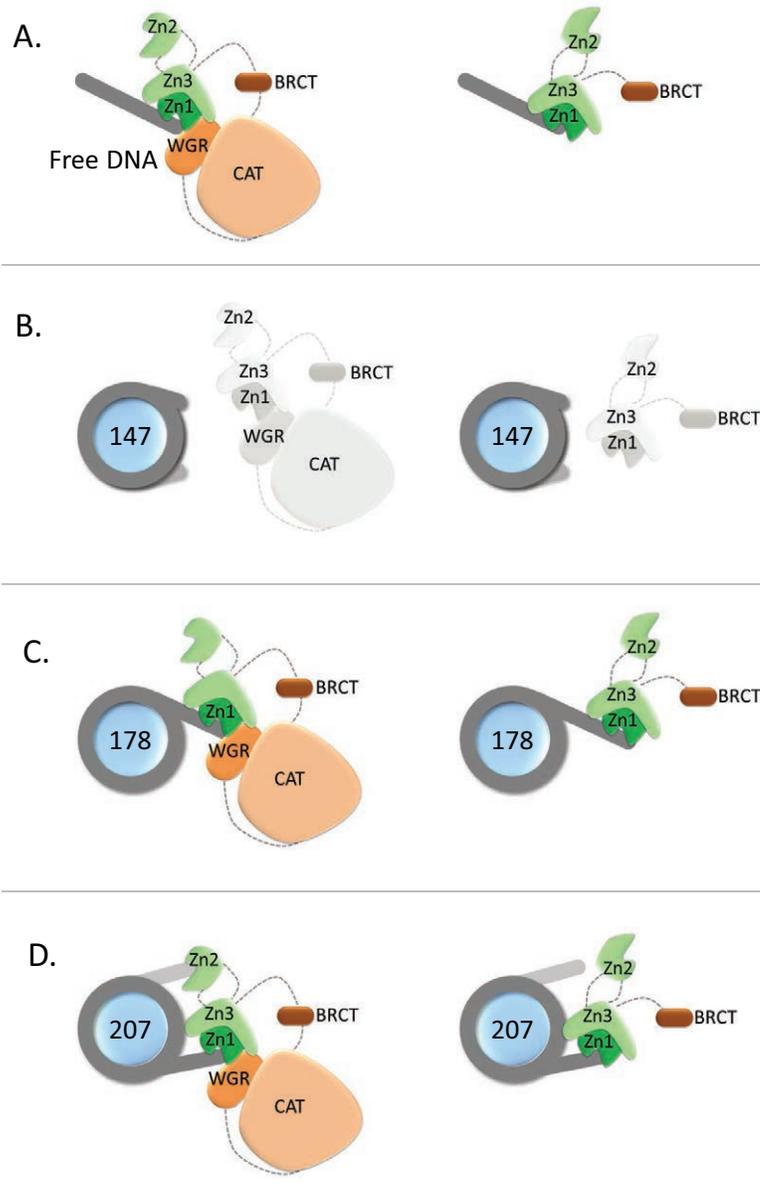


Figure 2.7: Model for *N*-parp and PARP-1 interactions with DNA and nucleosomes. **A.** Interaction with free DNA, as shown in [49] (cartoon adapted from [150]). **B.** Nucleosomes without linker do not interact with either PARP-1 construct. **C.** Nucleosomes with one asymmetric linker arm likely interact similarly with PARP-1 as they do with free DNA, with some steric inhibition from the nucleosome, as indicated by the weaker binding affinities. **D.** Interaction with nucleosomes with two linker arms: Both linker arms are engaged in PARP-1 binding, with Zn2 binding the second DNA linker. Regions from *C*-parp are required for correct orientation of Zn2.

nucleosome substrates of full length PARP-1 compared to *N*-parp suggest that the engagement of the WGR domain (and perhaps the catalytic domain) is required for Zn² to position itself optimally for interaction with the second linker arm (**Figure 2.7**). This effect is not observed on nucleosomes with just one linker arm, or on free DNA. Both *N*-parp and full length PARP-1 bind less tightly to these nucleosomes than they bind to 30Link DNA, which has the same sequence as the linker extension in Nuc178, suggesting steric interference to binding by the nucleosome core. Some controversy exists over the stoichiometry of PARP-1 in solution [48, 97], and on free DNA [125, 136, 151]. Here we have shown that a single PARP-1 molecule binds per nucleosome, consistent with the idea that PARP-1 and linker histone H1 interact similarly with PARP-1 [123]. Using the same approach, we find that H1 binds Nuc207 with higher affinity than full length PARP-1 ([142]; and unpublished results).

PARP-1 activity is reportedly induced by DNA damage, chromatin and even isolated histones [110, 123]. Using highly pure recombinant PARP-1 and well-defined DNA and nucleosome substrates, Our data demonstrates that PARP-1 is able to recognize DNA double strand breaks in the context of chromatin and is potently activated, consistent with its role as a first responder to DNA damage in eukaryotic cells. The high affinity of PARP-1 to nucleosomes and its activation by DNA and nucleosomes explains how PARP-1 regulates chromatin structure, transcription and DNA repair pathways. Additionally, the requirement of NAD⁺ for PARP-1 activation implies that other pathways utilizing NAD⁺ will further regulate PARP-1 activity in the various cellular processes [152]. However, to understand if and how PARP-1 redistributes from undamaged chromatin to sites of DNA damage, we have to quantify the interactions of PARP-1 with complex chromatin structures and chromatin components in the absence of DNA damage.

Acknowledgements

Nicholas J Clark, Michael Kramer and Uma M Muthurajan all contributed equally to this manuscript (presented here as a chapter).

Chapter 3

Inferential structure determination of multi-domain proteins from small-angle X-ray scattering data

Chapter Overview

The work published in this chapter was initiated by Thomas Hamelryck PhD of the Bioinformatics Centre in the Department of Biology at the University of Copenhagen, Denmark. The vision of his group is to combine low-resolution SAXS data with high-resolution structures for proteins that have not been fully described atomistically. The intent was to obtain Small Angle X-ray Scattering (SAXS) data from researchers who perform biochemical and structural biology research on complex biological systems that contain regions of flexibility or disorder for which structural descriptions may otherwise be unobtainable. The experimental approach was mostly focused on *in silico* calculations that were meant to determine the three-dimensional structure of the average molecule in solution. In terms of contributions, this author contributed the SAXS data of the truncated N-terminal and full-length PARP-1 proteins used for the novel computational analysis. Additionally, as a means to better understand the results from the novel computational approach, this author worked closely with the lead author Kasper Stovgaard to perform the more readily accepted SAXS analysis routinely published in the literature. This chapter is presented in essentially the same form as a submission to the Journal PLoS Computational Biology. Although, the article was denied publication the referees noted that the approach showed promise but was not quite developed enough to portray full confidence in the findings. Our collaborators are addressing the reviewer's concerns, with a planned resubmission in another suitable journal.

Summary

Obtaining information on the structure of large, flexible proteins is notoriously difficult, yet of prime importance given their central role in medicine and biotechnology. Recently, the use of low resolution small-angle X-ray scattering (SAXS) data has made it possible to obtain structural information on proteins that consist of multiple domains connected by flexible linkers. Current methods require subjective choices, such as the size of the ensemble or how to reconcile the influence of the SAXS data with the requirements of viable protein geometry. Here, we present a novel method that is powered by a well-justified probabilistic model, which takes the limited information content of SAXS data into account to avoid over-fitting. The method produces ensembles that consist of structures that closely fit the data and a geometry that is controlled by probabilistic models of local protein structure. We show that even such “conservative” ensembles contain a surprisingly large conformational variability. The method provides useful results if the shape of the ensemble is fairly close to the shape of the bulk of the individual conformations in solution. We carefully validated the method using artificial data and previously published experimental data on the cellulase Cel5G and a truncated form of PARP-1, a DNA binding protein that consists of six domains and is a potential target in cancer therapy. Subsequently, the method was applied to novel, unpublished data on full-length PARP-1, illustrating that biologically meaningful results can be obtained from large, flexible macromolecular systems of significant biological relevance.

Proteins are vital to the function of most cellular functions and are the principle target for drugs. While the three-dimensional structure of proteins is essential for their function, it remains tedious and costly to obtain high-resolution protein structures. In particular, high-resolution structure determination methods have limitations with regards to flexible, multi-domain proteins. This group of proteins is of special interest, as they are involved in critical functions such as regulation of gene expression and the cell cycle. Small-angle X-ray scattering (SAXS) is a method that provides easy access to low-resolution structural information, even for flexible, multi-domain proteins. Alone, it is insufficient for detailed

structural analysis. We present a method for inferential structure determination that combines SAXS data with a probabilistic model of protein structure. Our method provides an ensemble of structures that are all in agreement with the SAXS data. The method was successfully applied to two previously studied proteins with known high-resolution domains connected by flexible linkers. We applied the method to the challenging six domain protein PARP-1 using unpublished SAXS data, producing biologically relevant information on the relative domain arrangements.

Introduction

Despite advances in protein structure prediction, it is still not routinely possible to reliably fold proteins *in silico* [153]. However, easily obtainable low-resolution experimental data can provide valuable complementary constraints for structure prediction and determination. Small-angle X-ray scattering (SAXS) is a well-established low-resolution method for structure determination [154, 155]. Methods for high-throughput automated SAXS analysis of proteins have recently been developed [156, 157], potentially facilitating a large increase in available SAXS data. SAXS curves are most commonly collected up to a maximum scattering vector of 0.3–0.6 Å⁻¹ corresponding to a spatial maximum resolution of 10 to 35 Å. This makes SAXS a valuable constraint for investigating the flexibility of larger multi-domain proteins. Most eukaryotic proteins are indeed composed of multiple conserved domains connected by flexible linkers [158, 159]. Such proteins are of particular biological interest as they carry out vital functions in regulation of gene expression, cell growth, cell cycle, metabolic pathways, signal transduction, protein folding and transport [158]. Structural data on such proteins are few as they are notoriously difficult to study by high-resolution methods [154] and likely exist in multiple conformations. On the other hand, the atomic structures of the individual domains are often known. This work therefore focuses on obtaining information on relative domain orientations in multi-domain proteins for which only the structures of the individual domains are known.

Current methods for modeling flexible, multi-domain proteins can be divided into two groups, dependent on whether they fit single structures to the SAXS curve or employ ensembles of structures. The most widely used programs from the first group are the rigid body modeling tools in the ATSAS package [160], including SASREF and BUNCH. These tools are based on spherical harmonics algorithms and refine models against experimental SAXS curves using simulated annealing protocols. SAS-REF can be used to model the quaternary structure of protein complexes using random translations/rotations of the subunits in a grid search. BUNCH can model multi-domain proteins linked by flexible loops. The linkers are represented as interconnected chains, described by one scattering body per residue, and are sampled by random rotations about the axis connecting two such dummy residues. BUNCH can be used to fill in unknown linker fragments between domains of known structure; the resulting model reflects a single average conformation of the protein. Recently, methods utilizing ensembles of structures to fit the experimental SAXS data have been developed. Bernardo *et al.* [161] use a genetic algorithm to select a number of different conformers for flexible, multi-domain proteins to enable the average theoretical scattering curve to fit the experimental data. The structures in the initial pool – typically 10,000 – are randomly generated from $C\alpha$ models. The resulting number of conformers in the ensemble and their diversity reflects the flexibility of the system. A related method uses molecular dynamics simulations to investigate the conformational space of the structure, followed by a selection of a minimal ensemble by a genetic algorithm [162].

We recently proposed a computationally efficient method for the calculation of SAXS profiles from protein structures [163]. The method is based on the Debye formula using a set of coarse-grained form factors, and proved successful in a decoy recognition experiment.

Materials and Methods

Protein expression and purification

Human Parp1-486 was cloned and expressed as described in [164]. Full-length human PARP-1 was expressed in *Sf9* insect cells according to [92]. Cell lysates were thawed from -80 C and sonicated (3 x 5 s, output 6.5, duty cycle 65%, Branson sonifer 450) on ice in ice-cold lysis buffer (300mM NaCl, 25mM Tris/HCl, pH 8, 1mM β -mercaptoethanol and 1mM PMSF). Cell lysates were then cleared by centrifugation (14,000 rpm for 30 minutes at 4°C). DNA was removed by addition of 1.0 mg/ml salmon sperm protamine sulfate (Sigma-Aldrich) followed by centrifugation (14,000rpm for 30 minutes at 4°C). The supernatant was then precipitated by a two- step ammonium sulfate treatment. The first precipitation was from 0-30% (164g ammonium sulfate solid was added to 1000ml to give a 30% saturation) and the second step was from 30-70% (249g ammonium sulfate solid was added to 1000ml to give a 70% saturation). The ammonium sulfate precipitations were performed at 4°C while stirring overnight. The precipitated proteins were resuspended in heparin chromatography buffer A (100mM NaCl, 25mM Tris/HCl, pH 8 and 1mM β -mercaptoethanol) and loaded onto a HiTrap Heparin column (GE Healthcare) and eluted with a linear gradient of heparin chromatography buffer B (1.5M NaCl, 25mM Tris/HCl, pH 8 and 1mM β -mercaptoethanol). Further purification included size-exclusion chromatography and cation exchange using a HiTrap SP ff column (GE Healthcare).

Light Scattering

Samples (2mg/ml) of either Parp1-486 or full-length PARP-1 in 50mM Tris/HCl, pH 8, 300mM NaCl and 1mM tris(carboxyethyl)phosphine (TCEP) were loaded onto an ÄKTA purifier HPLC system. A Superdex S200 (10/30) size-exclusion column (GE Healthcare) was used to characterize the proteins. The proteins were continuously run at a flow rate of 0.3ml/minute from the S200 column into an inline Dawn Heleos II (Wyatt Technologies) multi-angle light-scattering instrument, immediately followed by a REx refractive index detector (Wyatt Technologies). The proteins were further analyzed using a differential index of refraction (dn/dc) of 0.185.

Small Angle Scattering

SAXS measurements were performed at two separate beam lines. Data for Parp1-486 as well as the full-length PARP-1 were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) at the beam line 4-2. Samples were analyzed at room temperature using a Mar-CCD165 detector. The data images were normalized, the sample transmissions were corrected, if multiple images were found not to have time dependent changes in scattering, those images were combined prior to channel- by-channel buffer subtraction. Intensity curves for the Parp1-486 and full-length PARP-1 were both collected at concentrations ranging from 1-3 mg/ml. Scattering was independent of protein concentration indicating that inter-particle effects such as repulsion or aggregation were negligible (data not shown). Samples of full-length PARP-1 were also measured at Berkeley's SIBYLS beam line (12.3.1) at the Advanced Light Source (ALS) using a Mar CCD area detector. The data images were circularly integrated, normalized and the buffer contribution subtracted. Intensity curves for 3.8, 1.9 mg/ml and 1.3 mg/ml concentrations showed no concentration dependent inter-particle effects (data not shown).

Initial data processing was performed using the program PRIMUS [165] and the radius of gyration calculated using the Guinier approximation as implemented in the GNOM program [166]. The DAMMIN program [167] was used to generate the envelope models presented in **Figures 3.5** and **3.7**. Ten independent dummy-atom models were calculated from the scattering data for each PARP-1 construct (in slow mode) and superimposed using SUPCOMB [168]. The resulting models were averaged and filtered using DAMAVER [169] and converted to volumetric maps using the program pdb2vol in the SITUS suite [170].

Input domain structures

For the study with artificial data, the PDB file 1CCZ was used. The Cel5G domains were modeled using high-resolution structures 1AIW and 1EGZ for the cellulose binding and catalytic domains

respectively. For Parp1- 486 the following PDB files were used as input for the Zn1, Zn2, Zn3 and BRCT domains: 2DMJ, 2CS2, 2RIQ, 2COK. These were also used for the full-length PARP-1 modeling with the addition of 2CR9 for the WGR and 2RCW for the catalytic domain.

Probabilistic model of the side chain scattering bodies

The calculation of the SAXS likelihood requires coarse-grained structures; the side chains need only be represented by a single scattering body placed at their center-of-mass (CoM) [163]. If needed, full-atom side chains can later be added, for example using BASILISK [171] or IRECS [172]. The CoM position is calculated from all heavy atoms in the side chain; we excluded the β -carbon since its position is approximately fixed for a given backbone conformation.

We defined the position of the CoM by defining a local reference frame based on the residue's backbone atoms (see **Figure 3.9**). The X-axis is defined by the N \rightarrow C vector. The vector that defines the Y axis is perpendicular to this vector, and lies in the N – C α – C plane. The Z-axis is defined by a vector that is perpendicular to both vectors, resulting in a right hand coordinate system.

Bayesian networks are graphical models, where the nodes represent random variables and the edges encode their conditional independencies [173]. We formulated the probabilistic model of the CoM as a simple Bayesian network, consisting of a single hidden node, and four observed nodes that concern the variables of interest used in sampling (**Figure 10**). The model is named COMPAS, which loosely stands for “center of mass placement of side chain centers”. In the next two paragraphs, we describe the four observed nodes in COMPAS.

Side chain conformations have a strong correlation with the conformation of the backbone [171]. In order to capture this relation, we included the two main degrees of freedom in the protein backbone, namely the ϕ and ψ angles, into the model. In order to model these angles we used the cosine variant of bivariate von Mises distribution, [174] which has successfully been applied in this context before [175]. This pair of angles is thus modeled by one bivariate von Mises node.

The amino acid type is represented by a discrete distribution that can adopt 18 states, corresponding to all amino acids except proline and glycine. The direction of the CoM was modeled using the 5-parameter Fisher-Bingham distribution [176, 177], which is a distribution over unit vectors. Analyzing the training data revealed that the placement of the CoM atom is limited to a small number of distinct distances for each amino acid. Therefore, the CoM distance from the C α atom is described by a discrete distribution of three states. Each state corresponds to a certain distance, which also depends on the amino acid type and the hidden node can adopt 120 states. The distance parameters are included in **Table S3.15**.

COMPAS is used in the following way. First, the amino acid type and the values of the ϕ and ψ angles values are specified using the corresponding nodes. Then, a value is sampled for the hidden node, conditional upon this information. Finally, a unit vector and distance are sampled conditional upon the sampled hidden node value. The unit vector and distance uniquely specify the position of the CoM.

The COMPAS model was trained on angles derived from 1703 high-resolution crystal structures that were culled from the Protein Databank using the Pisces server [178]. All structures selected have been solved to a resolution of at least 1.6 Å, an R-value of 0.25 or better and a pairwise sequence similarity of 25% or less. Subsequently, the structures were processed using the REDUCE software [179], correcting common misconfigurations such as flipped histidine rings. From this initial set around 10% randomly selected structures (171 structures with 31,229 observations) were retained as a test and validation set. The remaining 1532 structure or 277,975 observations were used to train the model. The COMPAS model was implemented and trained using our freely available Mocapy++ toolkit [180]. The parameters in the model were estimated using the stochastic expectation maximization (EM) algorithm [181].

MCMC sampling

Starting from the initial structures, MCMC sampling is performed using our molecular modeling software PHAISTOS [182]. We applied a generalized ensemble with a weighting according to the inverse cumulative density of states [183], called the 1/k-ensemble, to obtain exhaustive sampling in the high

probability regions. A non-uniform, adaptive binning of the energy space was used to ensure efficient sampling for both smaller and larger proteins. Samples were dumped every 5000 iterations, typically resulting in thousands of random samples. The minimum energy structures were saved for further analysis as well. Only the linker regions were changed during sampling; the domains themselves were kept fixed.

Various moves were investigated for this sampling. The best results, in terms of number of sampling iterations needed to reach the lowest energy structures, were achieved with a combination of two kinds of moves. So-called “pivot” moves, which are using the TorusDBN prior [175] and updating 1-5 pairs of dihedral angles at a time, and the more contained “local pivot” moves, also including the TorusDBN prior, and restricted to change one dihedral angle at a time with a smaller variation. For the side chain CoM, a new position was sampled according to the COMPAS model given the amino acid label and the backbone conformation at this residue. For the refinement run, all the heavy atoms of the side chains were inserted, and the dihedral angles were sampled using the probabilistic model BASILISK [171].

Based on a diverse validation set of proteins [163], we expect the sampled energy to reach reasonably low regions for a run, when the corresponding χ^2 values are around 0.5–1.0 using the following formulation;

$$\chi^2 = \frac{1}{Q-1} \sum_a ((I_q - I'_q) / \sigma_q)^2$$

with Q the number of q-bins, I_q and I'_q the input and theoretically calculated scattering intensities and σ_q the estimated experimental error. Multi-threaded runs were applied for PARP-1 simulations due to the size of these constructs. For the latter, we ran three independent Markov chains, which share the entropy information but explore the space independently; this resulted in over 100 million iterations.

Energy evaluations

Following the Bayesian calculus, the posterior probability used to drive the sampling was defined

$$\begin{aligned} P(X|D, A) &\propto P(D|X, A)\pi(X|A) & (2) \\ &= \prod_{q=1}^Q \mathcal{N}(I_q|I'_q, \sigma_q)\pi(\bar{\phi}, \bar{\psi} | A) \end{aligned}$$

where X is the protein structure under investigation with backbone dihedral angles $\bar{\phi}$, $\bar{\psi}$ and amino acid sequence A ; I is the experimental scattering curve; I' is the theoretical scattering curve calculated from X ; $\mathcal{N}(\cdot)$ is the Gaussian distribution; and the product runs over all q -bins. We used an estimated standard deviation for the experimental error σ_q [11]. The sampling was primarily investigating low energy structures and in effect

$$P(X|D, A) \propto P_{1/k}(E)\pi(\bar{\phi}, \bar{\psi} | A) \quad (3)$$

where

$$E = -\log P(D|X, A) = -\log \left\{ \prod_{q=1}^Q \mathcal{N}(I_q|I'_q, \sigma_q) \right\}$$

denotes the energy and $P_{1/k}(E) \propto w_{1/k}(E)g(E)$ is the marginal distribution over the energy implied by the $1/k$ - weights, $w_{1/k}(E)$, and the density of states, $g(E)$.

Apart from the SAXS energy, a simple steric hindrance term was also included in order to prevent clashes between atoms in the structure. This term assign an infinite energy to structures with separations of less than 1 Å between hydrogen atoms and 1.5 Å for all other pairs of atoms. Otherwise this term is zero.

For the ensemble analysis we selected the conservative ensemble as the samples within the lowest energy bin in the adaptive binning scheme, thus having no distinguishable difference in goodness of fit to the SAXS data.

As an additional validation, the scattering profiles for the centroid structures, computed by our coarse-grained model, were compared to the ones produced by the programs CRY SOL [184] and FoXS [185].

Both methods use full atomic details in the calculation of the theoretical scattering curve. CRY SOL applies a spherical harmonics algorithm for these calculations while FoXS uses the Debye formula for all atoms. The three different methods were in good agreement and all within the experimental uncertainty of the input curve. Examples for Parp1-486 and Parp1-1014 are included in Figures **S3.16** and **S3.17**.

Clustering and centroid selection

Calculating all pairwise distances for a large set of samples of structures in RMSD space is cumbersome and time consuming. We therefore used a clustering method where the structures are represented using Gauss integral tuned (GIT) vectors developed by Røgen and coworkers [186]. Here the protein topology is described by a 31 dimensional vector rather than atomic positions, which allows for a fast evaluation of the Euclidean distance between two such vectors. These GIT distances have been shown to correspond well with the RMSD measure [186]. Using the GIT clustering approach, the result of a typical MCMC sampling, in the order of tens of thousands of samples, can be clustered within seconds. The k-means algorithm [187] was used in the clustering step with k fixed at ten clusters. The GIT vectors do not represent large proteins sufficiently well, so this clustering approach was applied only for the ICCZ and Cel5G proteins.

The PARP-1 structure was modeled as fixed domains connected by long flexible linkers, which results in samples with a high geometrical variability. For Parp1-1014 the scattering profiles was collected with a resolution up to $q = 0.3 \text{ \AA}^{-1}$, meaning information primarily on molecular mass, radius of gyration, volume and a partial description of the geometrical shape. Given our prior information, the

tertiary structure of the domains and the primary sequence of the linkers, more than a single structure can fit the experimental SAXS data.

To appreciate the geometrical distribution of the individual domains, the structures were mapped to a feature vector containing the individual distances between the residues sterically placed at the extremes of the component domains:

- Parp1-486: residues 59, 96, 166, 206, 244, 285, 408, 450
- Parp1-1014: residues 43, 80, 150, 190, 228, 342, 441, 466, 563, 580, 668, 722

Collecting these vectors for each sample, a feature matrix was constructed and a PCA describing the roto-translations of the domain pairs was carried out. The first three PCA dimensions account for 88% percent of the variability for Parp1-486 and 86% for Parp1-1014.

A distance matrix between the structures was also computed using these PCA components. The employed distance measures were the Euclidean distance, the Pearson's correlation, the Spearman's rank and the Kendall's tau measure. For the distance matrices, a k-medoids algorithm [188] returned the cluster centroid structures used in the Results section. All the distance measures above agree on the cluster centroids for a given value of k clusters.

Results

Overview of the methodology

For NMR data the use of Bayesian inference for protein structure determination from experimental data was introduced by Rieping et al. [189]. This approach avoids arbitrary, ad hoc choices concerning the structure determination procedure [190] and is therefore both theoretically and practically attractive. We use a similar approach, thus formulating an alternative to methods that combine physical energies with pseudo-energies that then bring in the experimental data [191-193].

Following the Bayesian probability calculus, we formulate a posterior distribution $P(X|D,A)$ in the following way:

$$P(X|D, A) \propto P(D|X, A)\pi(X|A) \quad (1)$$

where X is a protein conformation, D is the experimental data and A is the amino acid sequence. The likelihood $P(D|X, A)$ brings in the information from the experimental data. For the likelihood we assume a Gaussian model, which is based on our previously developed coarse-grained Debye method [11]. The prior distribution $\pi(X|A)$ brings in prior knowledge on protein structure. As the prior distribution addressing the main chain, we use TorusDBN, our previously developed probabilistic model of local protein structure [175]. We complement TorusDBN with COMPAS, a probabilistic model of coarse-grained side chain positions introduced here (see the Methods section). Both models are used as generative models to sample plausible protein structures consisting of the main chain in atomic detail, and the centers of mass of the side chains. The latter are necessary for the application of the coarse grained Debye method [11]. The posterior $P(X|D,A)$ is the probability of a structure X given the experimental data D and the protein sequence A .

Structures are sampled from the posterior distribution using a Markov chain Monte Carlo (MCMC) approach. Starting from randomly sampled linker regions and following domain orientations, the corresponding scattering curve is calculated. This curve is evaluated with respect to the experimental data through the likelihood. In each step of the MCMC algorithm, a new structure is proposed using the prior, $\pi(X|A)$, and accepted/rejected based on the likelihood function, $P(D|X, A)$. To overcome the deficiencies of the standard Metropolis-Hastings algorithm, the simulation was done using $1/k$ generalized ensemble weights [194]. In effect, samples are drawn from the modified distribution:

$$P^*(X | D,A) \propto P^{1/k}(E)\pi(X|A)$$

where $E = -\log P(D|X, A)$ denotes the “energy” and $P^{1/k}(E) \propto w^{1/k}(E)g(E)$ is the marginal distribution over the energy implied by the $1/k$ -weights, $w^{1/k}(E)$, and the density of states, $g(E)$. Effectively, this approach ensures that the SAXS curves of the sampled structures are predominantly close to the experimental data, while maintaining a good local structure.

Our coarse-grained method does not account for the so-called hydration layer, which is a layer with a slightly higher density than the bulk water that surrounds a protein in solution. However, this effect is considered to be small at a resolution below 0.3 \AA^{-1} compared to the experimental uncertainties of the SAXS data [191].

Our method was applied to three different flexible, multi-domain proteins, as described below. In all cases, the high-resolution structure of each domain in the form of Protein Data Bank (PDB) files [195] and the amino acid sequence of the entire protein were used as input. A random, clash-free conformation of each linker region was sampled using TorusDBN, and used to attach the domains to each other. The resulting structure was then utilized as the starting point for the MCMC sampling. After convergence, the resulting structures are clustered, ranked according to the density of the clusters and the cluster centroids are analyzed in detail.

Artificial data

As a first test, our method was applied to a system with a known high-resolution structure, namely the two CD2-binding domains of CD58 (PDB code 1CCZ). Artificial SAXS data was generated from the native structure up to a scattering vector of 0.5 \AA^{-1} using the program CRY SOL [184] (see Methods). The scattering vector is defined as $q = 4\pi \sin \theta/\lambda$, where λ is the wavelength and θ is half the scattering angle. The resolution limit was chosen as a current typical experimental high-resolution limit. Choosing a protein with a known high-resolution structure allows for the evaluation of the sampled structures in terms of their root-mean-square deviation (RMSD) from the native structure. For this evaluation, two parts of the structure were designated as fixed domains, while the central linker was assumed to be flexible. The two fixed parts comprise residues 1 to 95, and 100 to 171.

Starting from a random conformation for the linker, MCMC sampling was carried out for 50 million iterations to ensure that convergence had been reached. The resulting samples were divided into ten clusters using k- means clustering (see Methods). The correlation of the RMSD versus the energy of

the samples shows that the lowest energy samples are all below 2 Å RMSD to the native structure (**Figure S3.1**). The cluster centroids and their calculated scattering curves are in good agreement with the SAXS data (**Figure S3.2**).

The resulting centroids are illustrated in **Figure 3.1**; centroids one to four are all very similar to the native structure (**Table 3.1**). In conclusion, the results from the experiment with artificial data confirm that the method works satisfactory in this case. We now turn to a more demanding application.

Cel5G

The cold-adapted cellulase Cel5G (EC 3.2.1.4) from the Antarctic bacterium *Pseudoalteromonas haloplanktis* consists of a catalytic module connected to a carbohydrate binding domain by an unusually long and flexible linker region. The importance of the linker region structure for the catalytic efficiency at low and moderate temperatures has been demonstrated by mutational studies followed by structural analysis using SAXS by Sonan and co-workers [196]. For the selected mutant, the three disulfide bridges in the linker were all disrupted, resulting in a reduction of the linker from 109 to 41 residues.

We modeled the structure of this Cel5G mutant using experimental SAXS data [196] as well as the sequence of the linker along with the known high-resolution structures of the catalytic module (293 residues) and carbohydrate-binding domain (60 residues) (see Methods). The experimental scattering curve was provided for a scattering vector q in the range of [0.022, 0.439] Å⁻¹ and was discretized in bins of 0.015 Å⁻¹ width for evaluations using the coarse grained Debye method.

MCMC sampling was carried out for twenty million iterations, clustering the resulting samples. The clusters were ranked according to their cluster density, which approximately coincides with the cluster centroid χ^2 values as defined in the Methods section (**Figure S3.3**). This is expected when the sampling is performed using the 1/k-weights, which favor structures with a high likelihood. Subsequently, the centroid structures were refined using the probabilistic model BASILISK [171] (see Methods), which allows filling in the atomic details of the side chains. This significantly improved the fit to the experimental data (**Table 3.2** and **Figure S3.4**), without affecting the domain distances and orientations;

the rearrangements were limited to internal changes in the linker region and movements of the side chains in this region.

The structures of the five best centroids are shown superimposed in **Figure 3.2**. Despite significant differences between the linkers of these structures – the RMSD is up to 4.6 Å between entire structures – they show orientations and domain-domain separations similar to the previously published model [196] obtained using the same data set with the program GASBOR [197]. For all centroids, the linkers are similar in their overall shape. The first residues of the linker region (residues 294–298) are extended, while the central part of the linker (residues 295–329) is more compact, possibly including some turn/helical elements, and generally displaying more variation. Lastly, residues 330 to 335 are more extended and show more variation than the first part. **Figure 3.3** shows a histogram of the inter-domain distances for the sampled structures; the inter-domain distance is defined as the Euclidean distance between the center-of-mass for the two domains (catalytic module and carbohydrate binding domain). We observe a quite narrow domain distribution with a peak around 77–80 Å. A similar MCMC run using the same input structures but without the SAXS constraint is shown for comparison. Here, the distribution of inter-domain distances is broad, approximately ranging from 33 to 118 Å.

A histogram for the radius of gyration (R_g) calculated for each sampled structure is shown in **Figure 3.3**. Again, including the SAXS information leads to a narrow R_g distribution (between 32 and 36 Å) compared to an unrestrained simulation (between 22 and 47 Å). The lowest energy and first centroid structures have R_g s equal to 32.4 and 32.5 Å, respectively. This is in good agreement with the expected R_g calculated from the experimental data using the Guinier approximation, which is equal to 32.8 ± 0.9 Å. The average SAXS curve of the conservative ensemble for Cel5G, here represented by the over 6000 samples in the first energy bin (see Methods), is practically identical to the lowest χ^2 curve in the sampling (**Figure S3.5**). For Cel5G, the best centroid structures agree with previously published bead models, additionally providing information on the conformations of the linker. These conformations are similar in length but show some variation in the local structure.



Figure 3.1. 1CCZ centroid structures. Top: Best five centroid structures (c1 to c5) of 1CCZ compared to the native structure (green). The centroids are color-coded from dark blue (c1) through light blue (c5). Bottom: Next five centroid structures (c6 to c10) of 1CCZ compared to native structure (green). The centroids are color-coded from dark orange (c6) through light orange (c10). All centroids are superimposed to the first domain (left-hand side). The figure was prepared using PyMOL [22].

Table 3.1. 1CCZ cluster centroids with corresponding energies and the RMSD to the native structure. The energy is defined as the negative logarithm of the likelihood.

| Centroid # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Energy | 566.9 | 567.3 | 566.9 | 567.5 | 573.8 | 578.8 | 574.7 | 577.8 | 585.6 | 593.0 |
| RMSD [\AA] | 0.76 | 0.99 | 1.22 | 0.34 | 4.77 | 8.65 | 9.40 | 8.34 | 10.67 | 8.92 |

Truncated PARP-1

Poly(ADP-ribose) polymerase-1 (PARP-1) is a member of a large family of enzymes that share homology to the C-terminal catalytic domain [198]. This family of proteins is best known for their roles in DNA repair, chromatin and transcriptional regulation of a number of genes. PARP-1 is highly conserved in eukaryotic cells and is often responsible for the synthesis of poly(ADP-ribose) (PAR) [92]. The enzymatic activity of PARP-1 is enhanced 10-100 fold in response to DNA damage, while tightly regulated in the absence of genomic stress.

PARP-1 consists of six domains of which the first two, Zn1 and Zn2, are zinc-fingers involved in DNA binding [164]. The third domain Zn3 is a zinc-ribbon domain that is required for DNA-stimulated activation of PARP-1, but not directly involved in DNA binding [199]. The Zn3 domain is immediately followed by a breast cancer 1 protein (BRCA1) C-terminus (BRCT) homology domain, which is thought to function as a protein–protein interaction module. The last two C-terminal domains are the WGR domain, named after the most conserved central motif, and the catalytic domain.

We analyzed two different structures: a construct that was truncated after the BRCT domain (Parp1-486) [164] and the full-length protein (Parp1-1014). High-resolution structures are available for all six individual domains and were used in the modeling (see Methods).

For the Parp1-486 construct, MCMC sampling was performed for twenty million iterations, up to convergence. The predicted scattering curves for the lowest energy structure and the best five cluster centroids were in good agreement with the experimental data (**Figure 3.4**); the corresponding χ^2 values are in the range 0.95–0.97.

The distributions of the calculated radius of gyration R_g and maximum dimensions D_{max} for the Parp1-486 samples are quite narrow (Figure S6). The mean values are somewhat smaller than those calculated from the data itself by GNOM [166]. However, the best centroid structure is also in best agreement with the values obtained from GNOM (**Table 3.3**).

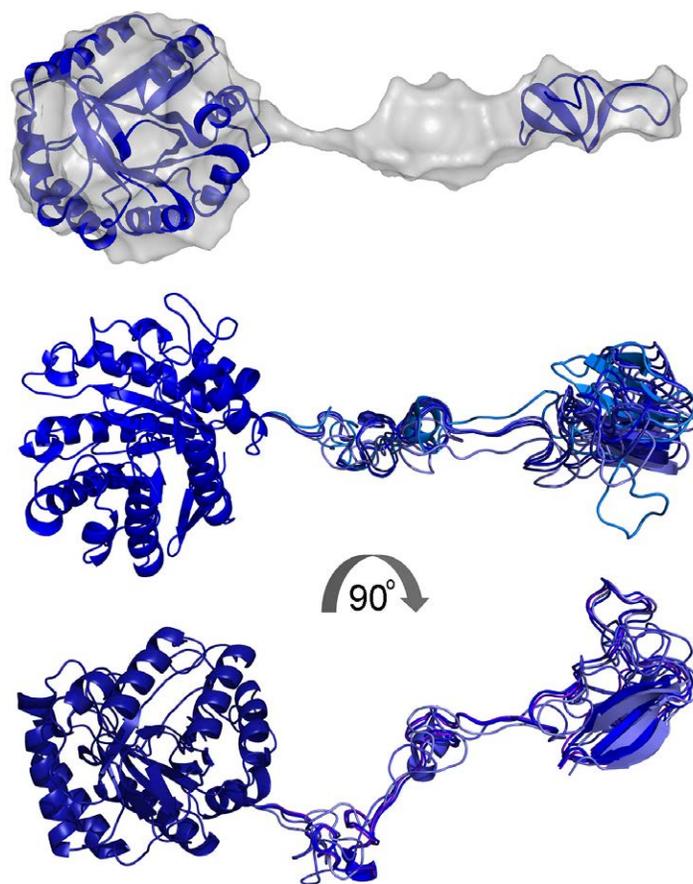


Figure 3.2. GASBOR model of Cel5G (grey) with overlay of domain crystal structures from Sonan et al. [23] (top) and superimposed centroid structures for our top 5 clusters in blue (center). The grey arrow indicates a 90 degree rotation around the horizontal axis. Figure prepared using PyMOL [22].

Table 3.2. The χ^2 -value, radius of gyration, Rg and inter-domain distance, D-D, for the refined Cel5G centroid structures.

| Centroid | χ^2 | Rg [Å] | D-D [Å] |
|----------|----------|--------|---------|
| 1 | 0.31 | 32.5 | 76.3 |
| 2 | 0.31 | 32.5 | 76.3 |
| 3 | 0.29 | 32.3 | 75.7 |
| 4 | 0.31 | 32.8 | 77.2 |
| 5 | 0.32 | 32.8 | 77.2 |
| 6 | 0.32 | 32.9 | 77.5 |
| 7 | 0.31 | 33.2 | 78.6 |
| 8 | 0.31 | 33.3 | 78.3 |
| 9 | 0.31 | 33.8 | 80.5 |
| 10 | 0.34 | 33.7 | 80.3 |

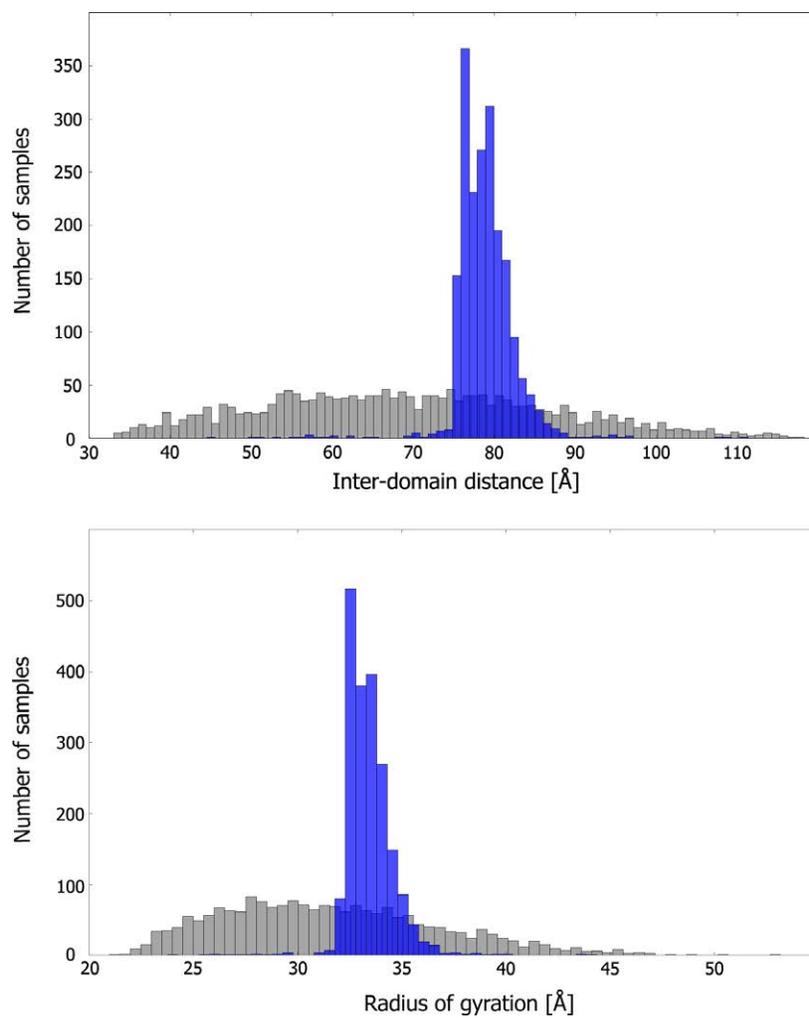


Figure 3.3. Cel5G inter-domain distance (top) and radius of gyration (bottom) for the sampled structures obtained with (blue) and without SAXS information (grey).

The structures of the centroids from the five clusters were investigated further. The curves and corresponding χ^2 values for these centroids are shown in **Figure 3.4**. The structure of the best centroid is S-shaped and elongated, in agreement with the DAMMIN model previously published for this data set [164] (see **Figure 3.5**).

The conservative ensemble indicates that very diverse structures are allowed by the experimental data. On the other hand, the average curve for the entire ensemble corresponds well to both the input data and the minimum χ^2 curve (**Figure S3.5**). Our results indicate that Parp1-486 is quite flexible in solution, but also has a preferred domain arrangement. In addition, the best centroid structure corresponds well to a previously published bead model for this construct [164].

Full-length PARP-1

No SAXS data has previously been published for the full-length PARP-1, but is presented here (see Methods). We found that Parp1-1014 behaves as a monodisperse monomer in solution similar to what was found for the truncated Parp1-486 [164] (**Figure S3.7**). Additionally, both proteins were free of concentration and scattering dependent inter-particle effects. SEC-MALS measurements show that the proteins have experimental molecular weights very close to the corresponding theoretical values. For Parp1-1014 the experimental weight was 111 KDa versus a theoretical weight of 113 KDa; the corresponding values for Parp1-486 were 57KDa and 56KDa, respectively. Although both constructs contain the BRCT domain, neither protein dimerizes under the conditions of the studies discussed here. For the Parp1-1014 modeling, sampling proved computationally challenging due to the sheer size and number of linkers involved. It was for this reason there was slightly worse sampling statistics than for the previously discussed proteins. Nonetheless, structures in very good agreement with the experimental input were sampled with χ^2 values in the range 0.35– 1.29 for the top five centroids (**Figure 3.6**). The R_g and maximum dimension distributions are also in reasonable agreement with the experimental expectations (**Figure S3.8**). The clustering method used for the other proteins is not suitable for structures of the size

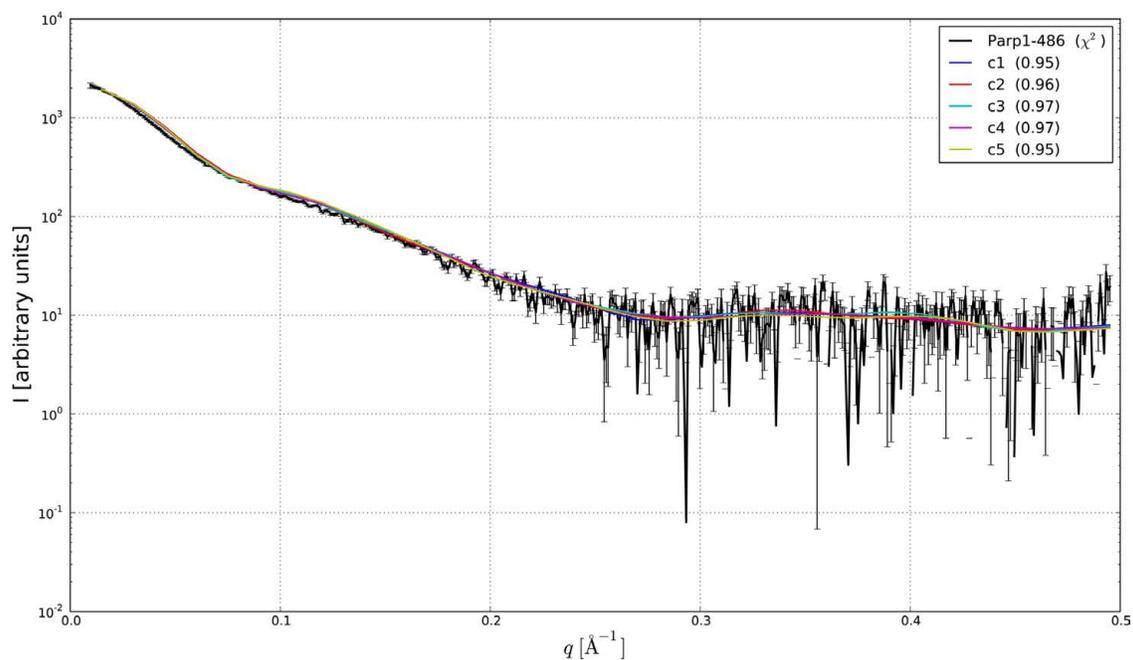


Figure 3.4. Experimental scattering curve for Parp1-486 with error bars and theoretical curves for the top five centroids (c1 to c5; see legend in the upper right corner).

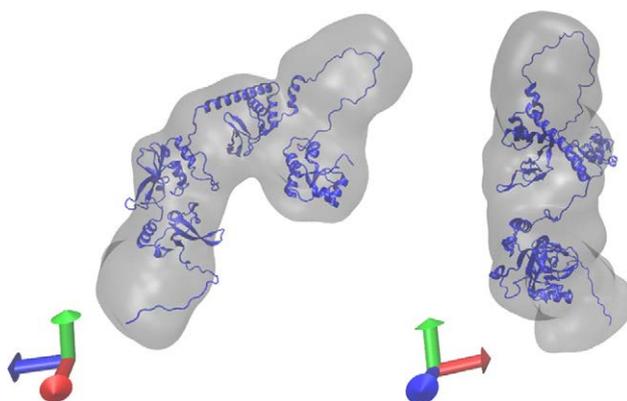


Figure 3.5. Comparison of the Parp1-486 top centroid structure (c1) in blue and an averaged DAMMIN model in grey, using the same input data (see Chapter 4 for comparison). The orientation is indicated by the axis on the bottom left-hand side of each model. Figure prepared using VMD [31].

Table 3.3. χ^2 , Rg and Dmax values for Parp1-486 centroid structures.

| Centroid | χ^2 | Rg [Å] | Dmax [Å] |
|---------------|----------|--------|----------|
| 1 | 0.95 | 40.5 | 155 |
| 2 | 0.96 | 38.9 | 136 |
| 3 | 0.97 | 39.0 | 123 |
| 4 | 0.97 | 39.8 | 145 |
| 5 | 0.95 | 38.6 | 112 |
| 6 | 0.98 | 39.2 | 115 |
| 7 | 1.14 | 39.2 | 121 |
| 8 | 8.23 | 56.9 | 133 |
| 9 | 0.98 | 38.4 | 166 |
| 10 | 1.17 | 38.0 | 113 |
| GNOM estimate | N/A | 44 | 165 |

of PARP-1. We therefore employed a PCA based clustering (see Methods), where cluster density does not correlate with χ^2 values (**Table 3.4**). Based on the χ^2 values, we consider the fourth centroid to be the most reliable of these structures, followed by the second and third centroid. Obtaining a complete structural alignment is difficult due to the protein size and the number and flexibility of the linkers. The centroid structures and their pair-distance distributions are illustrated in the supporting material (**Figures S3.9–S3.11**).

The fourth centroid is compared to an averaged DAMMIN model made using the same experimental data in **Figure 3.7**. Also in this case, the structural alignment is not straightforward, particularly since the DAMMIN model is almost ellipsoid and thus provides little information on the position of the domains.

The structure of the fourth centroid displays proximity of the Zn1, Zn2, Zn3 and WGR domains whereas the BRCT and PARP Catalytic (CD) domains are both isolated. The arrangement is different in the second centroid: the Zn1, Zn2 and the enzymatic domains are close together, as are the Zn3 and WGR domains. In the third centroid, the Zn1 and Zn2 domains are in contact but distant from the Zn3, WGR and CD domains. Again the BRCT domain is isolated, which is a common feature of the best three centroids. They also all agree on a tight connection between the Zn1 and Zn2 domains and between the Zn3 and WGR domains.

The curve for the average ensemble from structures in the lowest energy MCMC bin is close to the best χ^2 curve for the full-length PARP-1 (**Figure S3.5**). In order to analyze the main variability in the conservative ensemble, we conducted a structural principal component analysis (PCA) of these conformations. In this approach the eigenvectors for the covariance matrix of the ensemble structures are sorted by eigenvalues to determine the most important eigenmodes. The eigenmodes can then be visualized as structural displacements to describe the most significant variations. This is illustrated for the first three eigenmodes of Parp1-1014 in Movies S12–S14. The results show that the main variability is found in the distance between the Zn1-Zn2-Zn3 and CD domains, with observed

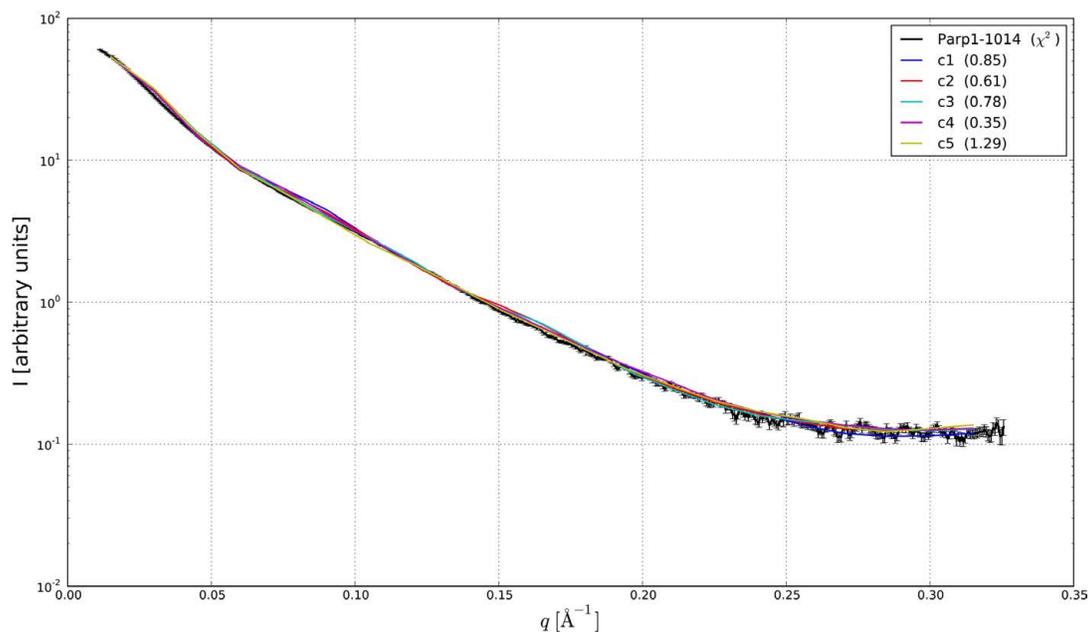


Figure 3.6. Experimental scattering curve for full-length PARP-1 and theoretical curves for the top 5 centroids (c1 upper right corner to c5; see legend in the).

Table 3.4. χ^2 , Rg and Dmax values for Parp1-1014 centroid structures.

| Centroid | χ^2 | Rg [Å] | Dmax [Å] |
|---------------|----------|--------|----------|
| 1 | 0.85 | 50 | 161 |
| 2 | 0.61 | 51 | 164 |
| 3 | 0.78 | 49 | 155 |
| 4 | 0.35 | 52 | 168 |
| 5 | 1.29 | 49 | 152 |
| GNOM estimate | N/A | 58 | 190 |

between the Zn3 and catalytic domains from around 15 to 25 Å. A smaller, hinge-like variation of the BRCT domain is also observed.

Next, we compare the domain orientation in Parp1-1014 and Parp1-486 for the shared domains. As superposition of diverse samples is problematic for large, flexible structures, we decided to analyze each pair of connected domains separately. Each domain and its neighbor in sequence with its connecting linker is shown for Parp1-486 and Parp1-1014 in **Figure 3.8**.

Figure 3.8 shows that the full-length construct appears to be more rigid than the Parp1-486 construct, especially for the short linker between the Zn1 and Zn2 domains where the Parp1-486 is highly flexible. Also, Parp1-1014 seems to have a few distinct, preferred orientations of the linker between the Zn3 and BRCT domains, whereas the Parp1-486 linker shows no such preference. One possible factor in the flexibility and elongation of Parp1-486 compared to full-length PARP-1, is that our Parp1-486 models lack a salt-bridge between Zn3 and the linker between Zn1 and Zn2. This salt-bridge is present in our models of Parp1-1014 (K105-D281).

Discussion

We have introduced a method for modeling flexible, multi-domain proteins. The local structure of the linkers is under the control of a probabilistic model of protein structure. Combined with a fast, yet accurate evaluation of the theoretical scattering curves of these samples this enables tens of millions of realistic structures to be generated in an MCMC run. Using experimental SAXS data as a constraint while sampling protein structures from a correct distribution of dihedral backbone angles and side-chain centers, we have shown that it is possible to reconstruct a two-domain protein with a short flexible linker to a precision below 2 Å RMSD from generated SAXS data. For a two-domain cellulase with a medium sized flexible linker our method obtains more diverse structures, which combined correspond well with a published model from the GASBOR program using the same input data. Additionally, our method samples the protein backbone in full detail and provides a plausible structural explanation for the shape

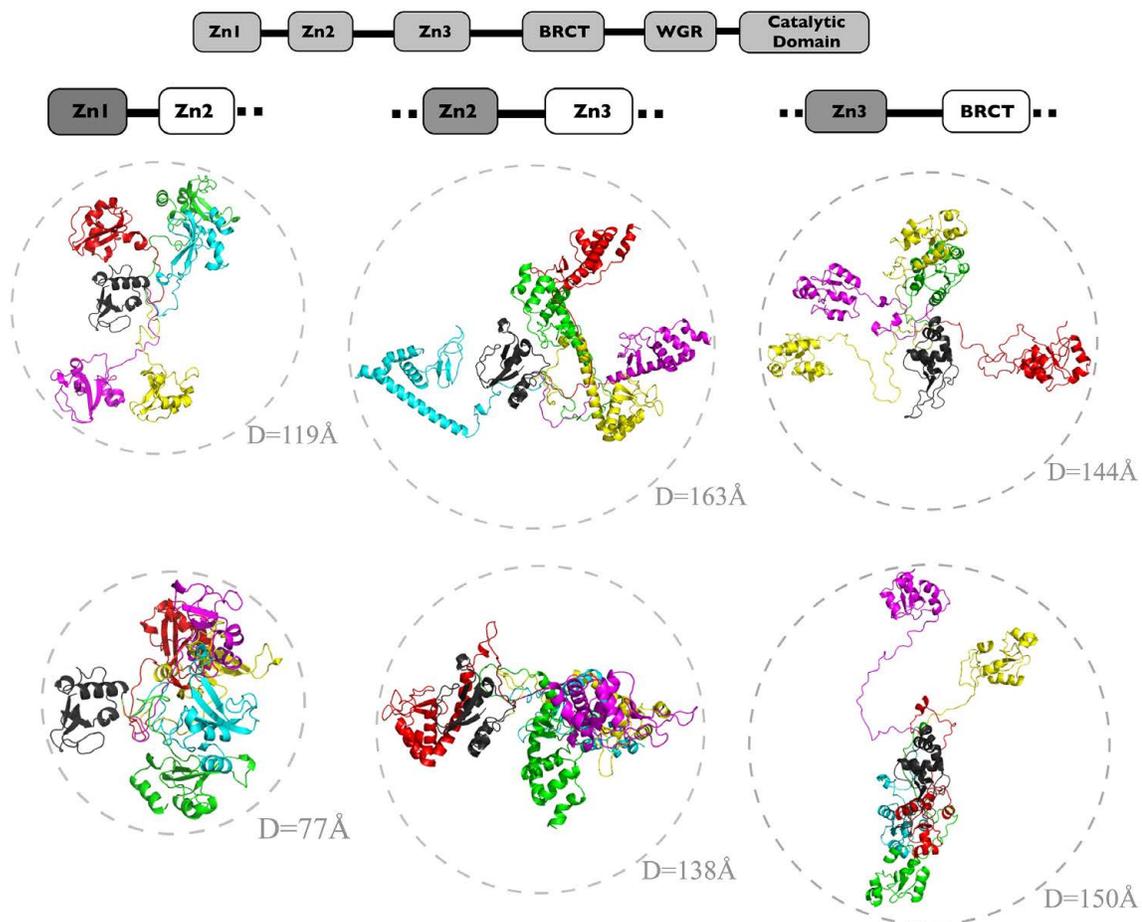


Figure 3.8. Variation in the linker regions. Relative domain positions for the top centroid structures of Parp1-1014 (top) and Parp1-486 (bottom) for domains Zn1-Zn2 (left), domains Zn2-Zn3 (center) and domains Zn3-BRCT (right). The first domain in each pair is superimposed for all centroids and is shown in grey while the second domains are colored in red (c1), cyan (c2), yellow (c3), green (c4) and magenta (c5). The approximate maximum dimension of each domain pair is outlined in grey. Figure prepared using PyMOL [22].

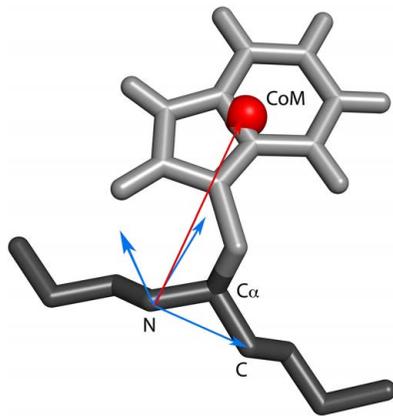


Figure 3.9. Center-of-mass placement. Center-of-mass vector for a tryptophan residue is shown as a red arrow. The axes in the local reference frame are shown in blue and the center-of-mass (CoM) used for the side chain scattering body position is marked as a red sphere.

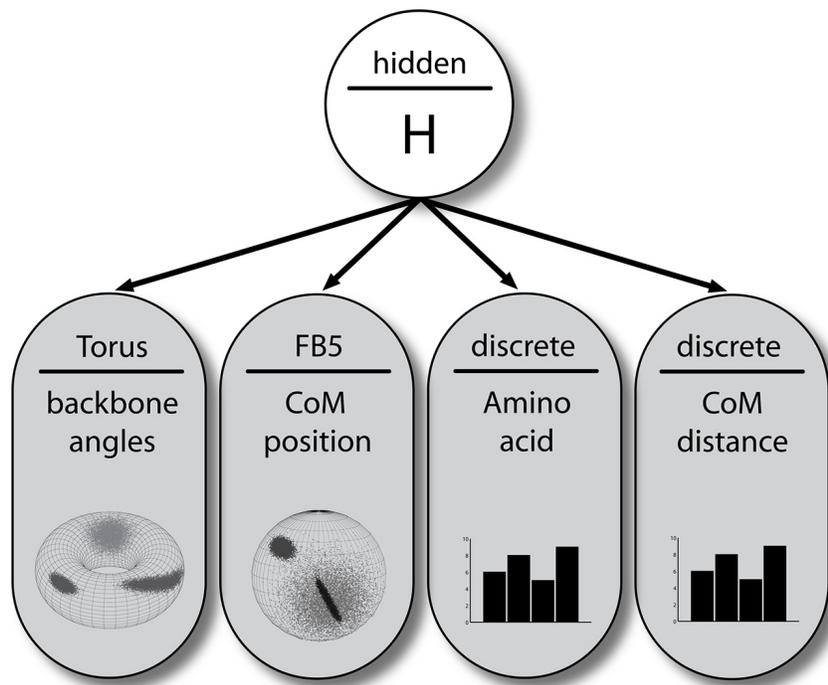


Figure 3.10. The COMPAS Bayesian Network model. The backbone angle is described by a bivariate von Mises distribution (on the torus) [13], the center of mass angular position with a 5 parameter Fisher-Bingham distribution (on the sphere) [44], the amino acid type with discrete node with 18 states, and the center of mass distance from the $C\alpha$ with a discrete distribution with three states. The hidden node, shown in white, can adopt 120 states. The observed nodes are displayed in grey.

and length of the linker region. Calculated values for the radius of gyration and maximum dimension concur with the expectations from the experimental data. The resulting S-shaped, elongated models of a truncated construct of the PARP-1 protein are likewise similar to a previously published DAMMIN model for the same data set. Both PARP-1 constructs proved quite flexible with different domain arrangements between the best centroids from our sampling. This result indicates that PARP-1 is a rather flexible protein, possibly continuously undergoing conformational changes in solution.

Currently our method is limited by the requirement that the global shape of the ensemble needs to be close to the shape of the individual structures. Input data from a flexible protein with several discrete conformations in solution will not be representable by our conservative ensemble and results in an MCMC run that stays in the high-energy regions with poor convergence. This limitation could be overcome by allowing multiple curve fitting in a dynamic ensemble scheme as used for instance in the Ensemble Optimization Method [9]. Nonetheless, our method potentially generates ensembles that can be regarded as representing a true ensemble in solution. To verify the ensemble aspect, we calculated an average curve from all theoretical scattering curves in the ensemble of all samples in the lowest energy MCMC bin (**supporting Figure S3.5**). In all four test cases, the average ensemble curve is effectively as good as the minimum energy curves sampled. In order to visualize the main variability in the ensemble, we conducted a structural PCA of the ensemble conformations for Parp1-1014 (Movies S12–S14). We observed a remarkable variety of structures in our conservative ensembles.

Since our method produces detailed conformations rather than bead or envelope models, the position and orientation of individual domains of the protein can be analyzed. This is of particular interest for eukaryotic multi-domain proteins where domain interactions and flexibility are vital for the function. One example is the observation that the BRCT domain in full-length PARP-1 is consistently found to be isolated and highly exposed compared to the flanking domains. The BRCT domain may not serve as a dimerization platform for the Parp1-1014 protein, since it exists as a monomer in solution, but its exposure may hold relevance in terms of a binding platform for known interaction partners like XRCC1 [200] or DNA ligase III-alpha [39].

Furthermore, the generalized ensemble MCMC sampling (see Methods) leads to a thorough investigation of the conformational space of the protein constrained by the SAXS data. This allows us to investigate the structural variability permitted by the data and the prior distribution through the variation in the obtained ensemble.

In conclusion, our method provides a novel solution with a firm theoretical underpinning to the inference of domain orientation in multi-domain proteins. The method is available as part of our PHAISTOS molecular modeling package [182] and can be obtained freely from Sourceforge1 under the GNU Library General Public License.

Acknowledgements

C. Andreatta, M. Borg, J. Frellsen and T. Harder are funded by the Danish Council for Strategic Research (NaBiIT, 2106-06-0009). W. Boomsma and K. Stovgaard are funded by the Danish Council for Independent Research (FNU, 272-08-0315 and FTP, 274-06-0380, respectively). We thank Sandro Bottaro and Jan Valentin for valuable discussions and suggestions on the parameters for the PHAISTOS and Muninn programs.

Contributions and Proposed Authorship for Declined Journal Submission;

✪Kasper Stovgaard^{1Y}, Christian Andreatta^{1Y}, Nicholas Clark^{2Y}, Tim Harder¹, Jesper Ferkinghoff-Borg³, Wouter Boomsma³, Mikael Borg¹, Jes Frellsen¹, Véronique Receveur-Bréchet⁴, Karolin Luger² and Thomas Hamelryck¹

Specific contribution descriptions: TH. Performed the computational experiments: KS CA. Performed the PARP-1 experiments: NC. Analyzed the data: KS CA NC. Contributed reagents/materials/analysis tools: TiH MB JeF JF WB KL VR. Wrote the paper: KS TH NC CA TiH.

✪ 1 -The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark 2 -Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado, USA
3 -Department of Biomedical Engineering, DTU Elektro, Technical University of Denmark, Lyngby, Denmark 4 IMR-CNRS - UPR3243, Marseille, France
Y -These authors contributed equally to this work

Chapter 3

Supporting Information

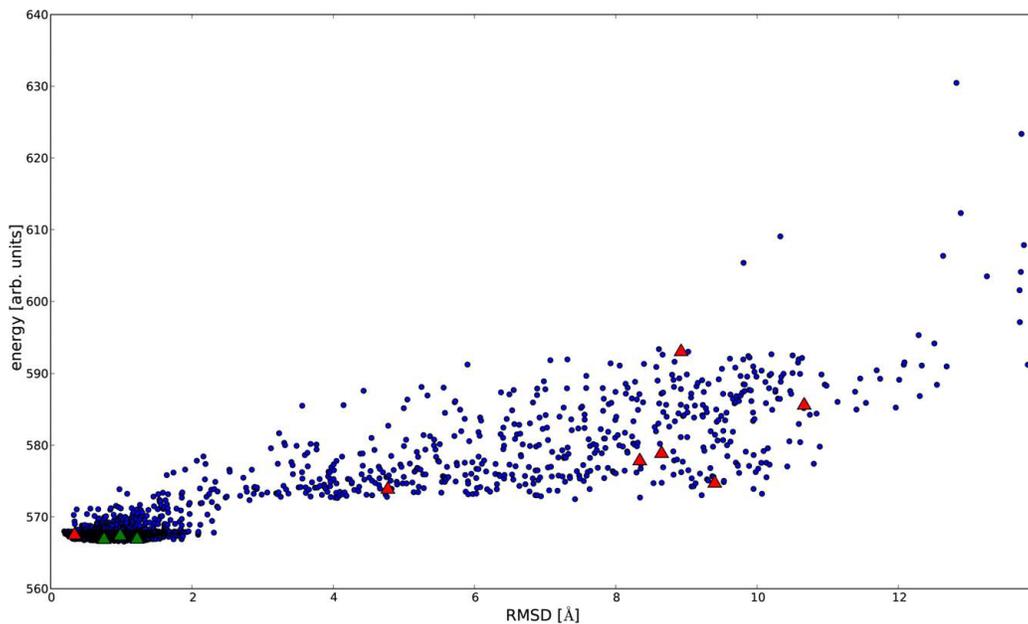


Figure S3.1. Correlation of energy (log-likelihood) and RMSD to the native structure for the 1CCZ samples. Maximum q in the input data was 0.5 \AA^{-1} . The cluster centroids are depicted as triangles with the top 3 centroids in green and the remaining in red.

Table S3.1. Modeling the distances for the CoM position. In the COMPAS probabilistic model of the CoM conformational space, we modeled the distance to the $C\alpha$ atom as a three-state discrete probability distribution. All values are given in units of \AA ; – represent states that were absent during the training phase.

| Amino acid | d_1 | d_2 | d_3 |
|---------------|-------|-------|-------|
| cysteine | 2.80 | – | – |
| aspartate | 2.92 | – | – |
| glutamate | 3.125 | 3.875 | – |
| phenylalanine | 3.79 | – | – |
| histidine | 3.57 | – | – |
| isoleucine | 2.25 | 2.70 | – |
| lysine | 3.80 | 4.20 | 4.6 |
| leucine | 3.05 | – | – |
| methionine | 3.185 | 3.85 | – |
| asparagine | 2.91 | – | – |
| proline | 2.29 | – | – |
| glutamine | 3.125 | 3.875 | – |
| arginine | 4.31 | 4.80 | 5.31 |
| serine | 2.43 | – | – |
| threonine | 2.17 | – | – |
| valine | 2.19 | – | – |
| tryptophan | 4.20 | 4.50 | – |
| tyrosine | 4.27 | – | – |

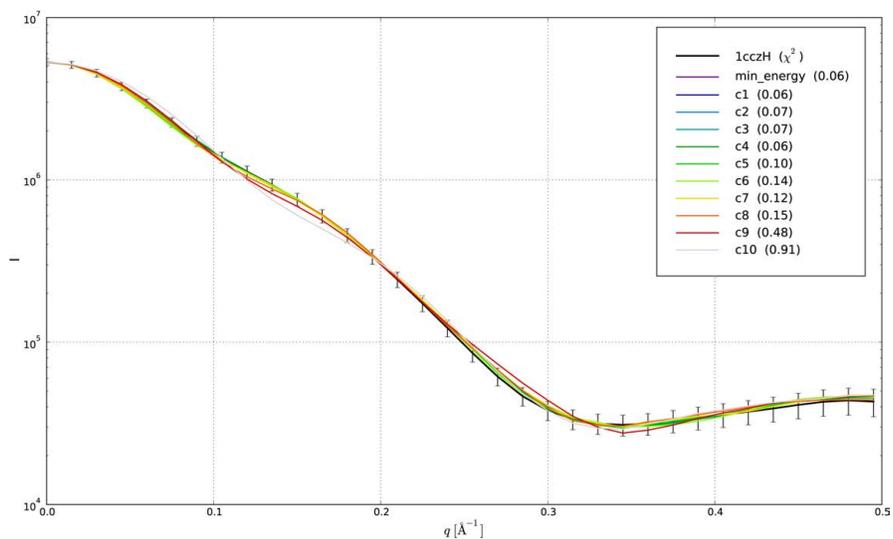


Figure S3.2. Generated SAXS curve for 1CCZ versus back-calculated scattering curves for the top centroids and minimum energy structure sampled (c1 to c10 and min_energy, see box in the upper right corner). Corresponding χ^2 values are included in parentheses. Error bars show the estimated experimental σ 's (see Methods).

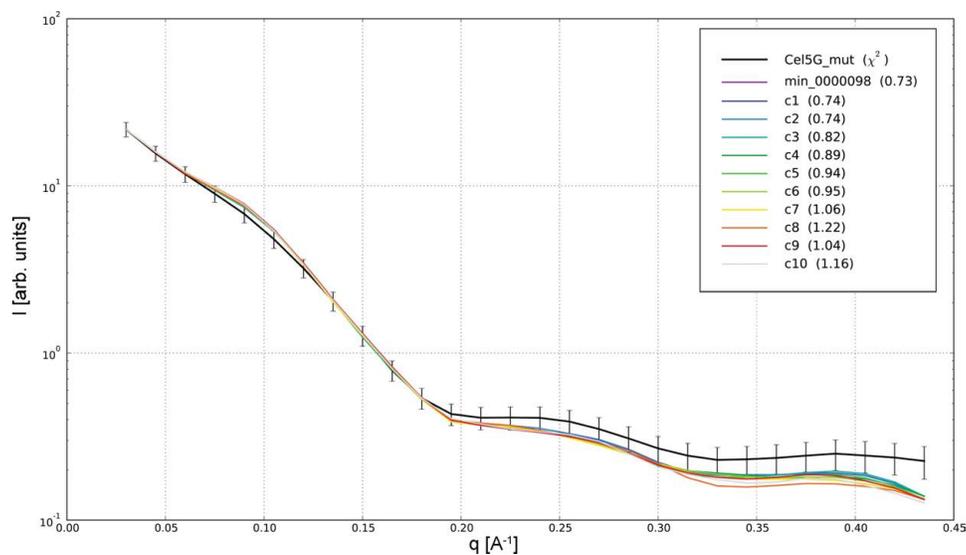


Figure S3.3. Theoretical scattering curves for the centroid structures compared the experimental input curve for Cel5G (Cel5G_mut and c1 to c10, see box in the upper right corner). The curve calculated for the minimum energy structure is also included.

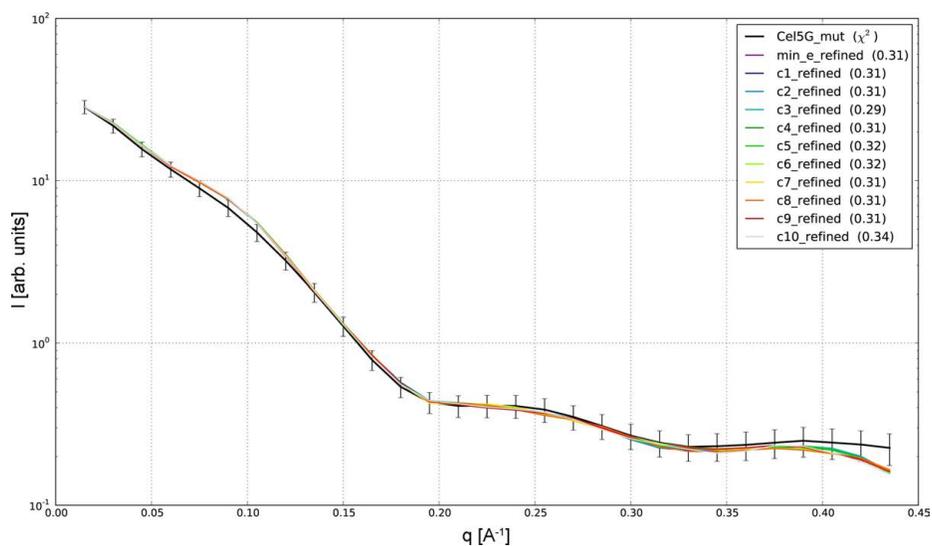


Figure S3.4. Theoretical scattering curves for the top centroid and minimum energy structures for Cel5G compared the experimental input curve after a full-atom refinement run (c1_refined to c10_refined and min_e_refined, see box in the upper right corner).

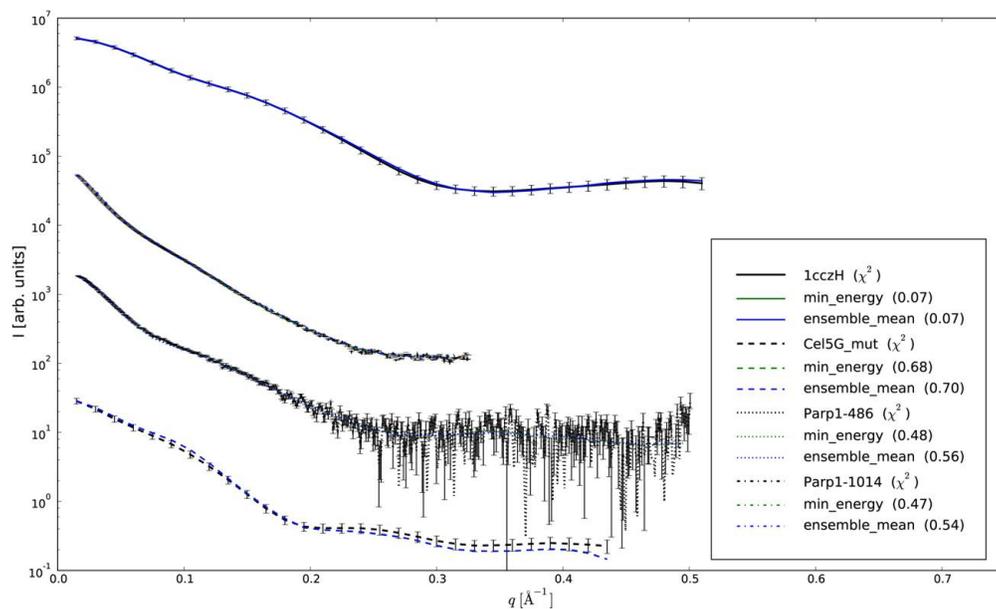


Figure S3.5. Minimum energy and the average scattering curves for the lowest energy bin ensemble with corresponding χ^2 values. From top to bottom: 1CCZ (735 samples), Parp1-1014 (3266 samples), Parp1-486 (3026 samples) and Cel5G (6646 samples), see box in the lower right corner.

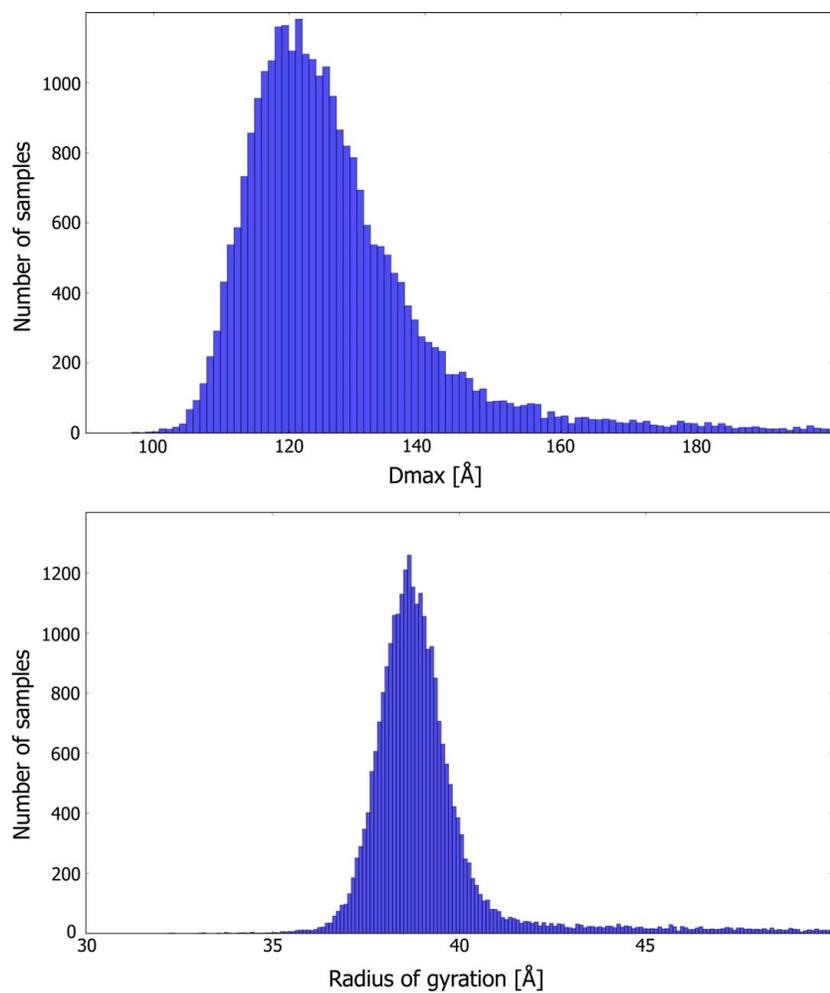


Figure S3.6. Distribution of the estimated maximum dimension (top) and radius of gyration (bottom) for the Parp1-486 samples.

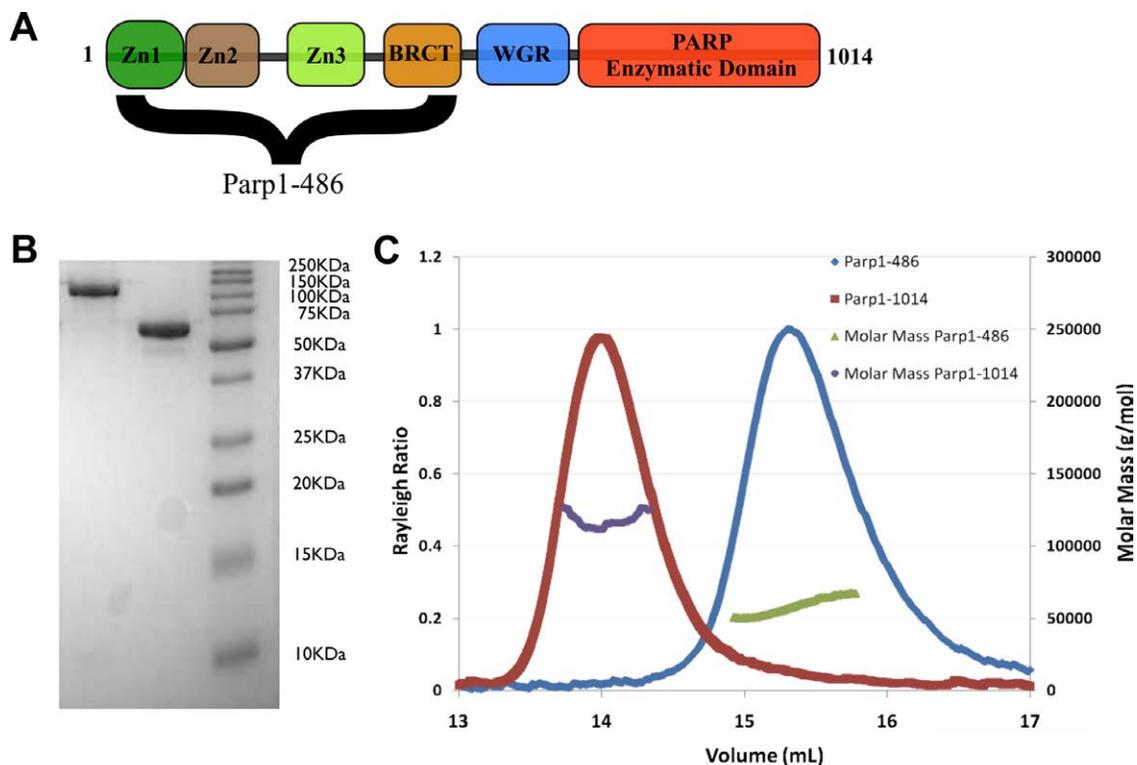


Figure S3.7. Laboratory studies of the enzyme Poly(ADP-ribose) Polymerase-1 (PARP-1). (A) The domain organization map of the full-length PARP-1 consists of six structurally independent domains all of which have been investigated as individuals in either NMR or crystallographic studies. The first four domains make up the Parp1-486 construct. (B) An image of a 12% SDS-PAGE gel containing samples of the full-length PARP-1 and the Parp1-486 proteins prior to SAXS experimentation. (C) Size-exclusion coupled multiangle light scattering elution profiles of the full-length PARP-1 and the Parp1-486 proteins. The static-light scattering derived average molecular weights are illustrated by red and blue dotted lines that bisect the Parp1-486 and full-length protein peaks, (56 kDa and 111 kDa) respectively.

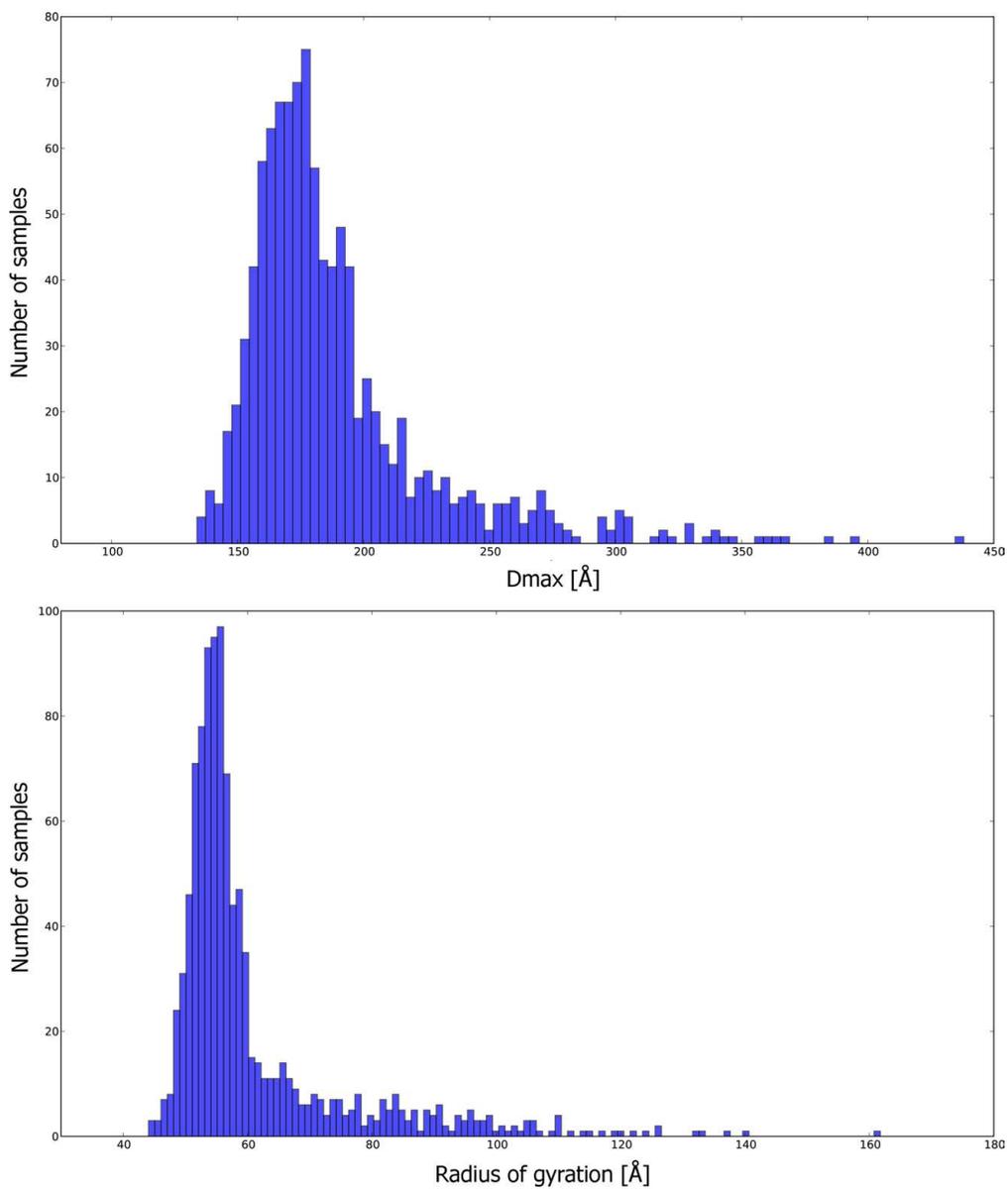


Figure S3.8: Distribution of the estimated maximum dimension (top) and radius of gyration (bottom) for Parp1-1014 samples.

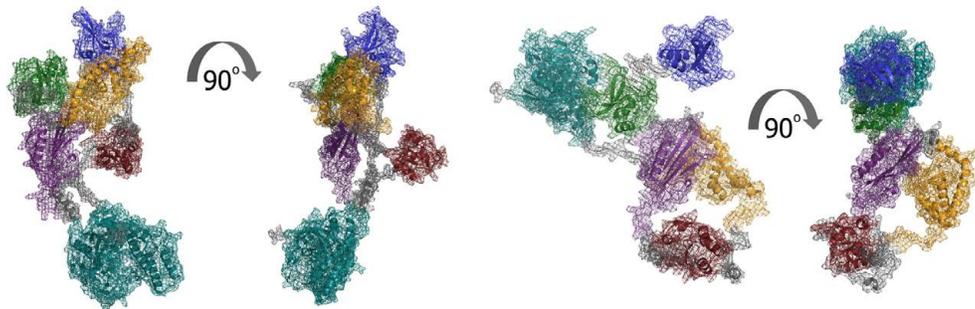


Figure S3.9: Full-length PARP-1 centroids 4 (left) and 2 (right) in two orientations. Domain colors are: Zn1 blue, Zn2 green, Zn3 yellow, BRCT red, WGR purple and PARP Catalytic domain cyan.

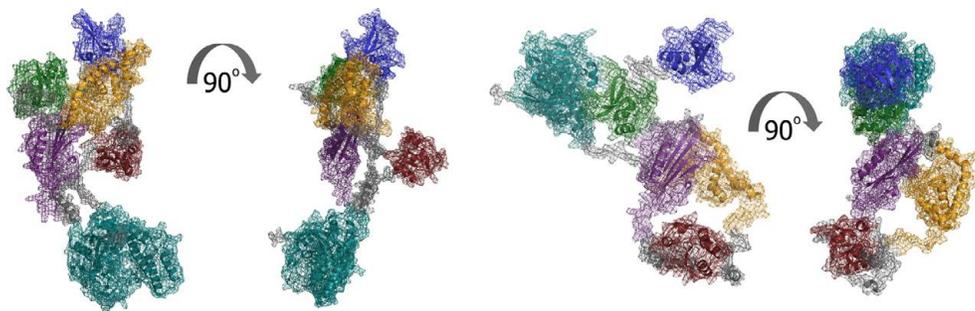


Figure S3.10: Full-length PARP-1 centroids 3 (left) and 1 (right) in two orientations. Domain colors are: Zn1 blue, Zn2 green, Zn3 yellow, BRCT red, WGR purple and PARP Catalytic domain cyan.

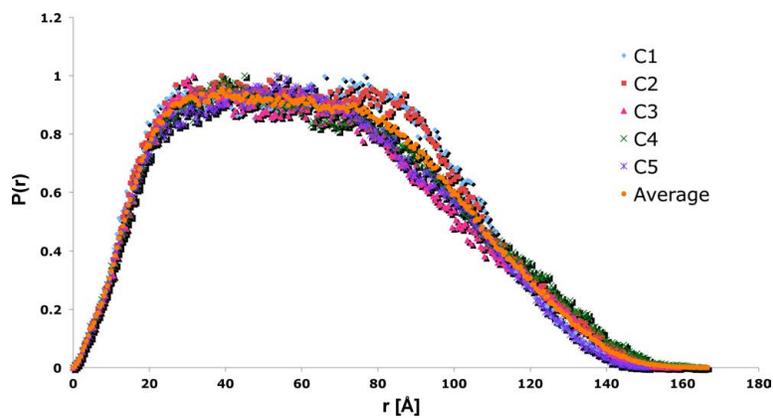


Figure S3.11: Pair-distance distribution functions. Estimates of the pair-distance distribution functions for each of the full-length PARP-1 centroids and their average function from the program GNOM.

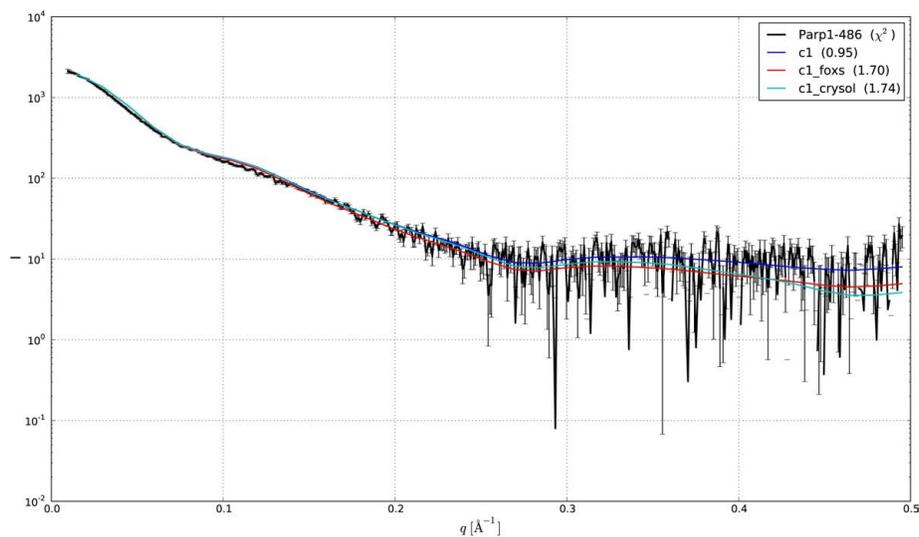


Figure S3.12: Theoretical scattering curves for the Parp1-486 c1 centroid using our coarse-grained model and the full-atom resolution CRY SOL and FoXS methods. Experimental input curve is shown for comparison in black and χ^2 values between the experimental and back-calculated curves are included (see legend, upper right corner).

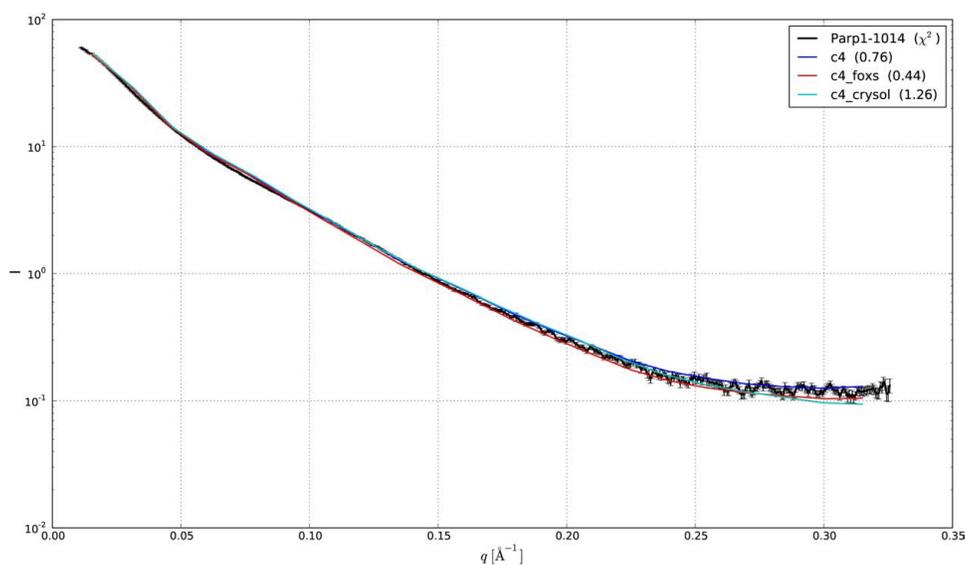


Figure S3.13: Theoretical scattering curves for the Parp1-1014 c4 centroid using our coarse-grained model and the full-atom resolution CRY SOL and FoXS methods. Experimental input curve is shown for comparison in black and χ^2 values between the experimental and back-calculated curves are included in the legend (upper right corner).

Chapter 4

Structural and Biophysical Studies of Human PARP-1 in Complex with Damaged DNA

Chapter Overview

This chapter is a peer reviewed journal article and it was accepted for publication in the Journal of Molecular Biology in 2010. Wayne Lilyestrom PhD designed and implemented the experimental direction of the research. I contributed my expertise in protein expression and high-quality protein purification. Additionally, I performed some of the biophysical analyses such as size-exclusion chromatography-coupled multi-angle light scattering and was in attendance for several of the late night SAXS data collection sessions at two synchrotron light sources. This chapter of the thesis served as a means to train me in the many structural biology and biochemical techniques that would later serve me during my own research on the full-length PARP-1 and its complexes with chromatin and DNA substrates (see chapters 3 and 5).

Summary

The enzyme Poly (ADP-Ribose) polymerase-1 (PARP-1) is a global monitor of chromatin structure and DNA damage repair. PARP-1 binds to nucleosomes and poly (ADP-ribosylates) histones and several chromatin-associated factors to expose specific DNA sequences to the cellular machinery involved in gene transcription and/or DNA damage repair. While these processes are critical to genomic stability, the molecular mechanisms of how DNA damage induces PARP-1 activation are poorly understood. We have used biochemical and thermodynamic measurements in conjunction with small angle X-ray scattering to determine the stoichiometry, affinity and overall structure of a human PARP-1 construct containing the entire DNA binding region, the zinc-ribbon domain and automodification

domains (residues 1-486). The interaction of this PARP-1 protein construct was evaluated with three different DNA damage models (DNA constructs containing a nick, a blunt end or a 3'-extension). Our data indicates that PARP-1 binds each DNA damage model as a monomer, with similar affinity, and in all cases results in robust activation of the catalytic domain. Using SAXS, we determined that the N-terminal half of PARP-1 behaves as an extended and flexible arrangement of individually folded domains in the absence of DNA. Upon binding DNA, PARP-1 undergoes a conformational change in the area surrounding the zinc-ribbon domain. These data support a model in which upon binding DNA, PARP-1 undergoes a conformational change to become an active nuclear enzyme.

Introduction

The enzyme Poly (ADP-Ribose) polymerase-1 (PARP-1) is a global monitor of chromatin structure and DNA damage. PARP-1 is a member of the large family of enzymes, which are defined by the PARP signature sequence, a 50-amino acid sequence within each member's enzymatic domain [23]. Several of the PARP proteins catalyze the cleavage of NAD⁺ into nicotinamide and ADP-ribose, using the latter to synthesize branched molecules of poly(ADP-ribose), a substrate that is covalently attached to a variety of nuclear proteins. In eukaryotes, members of the PARP family are best noted for their roles in DNA repair, chromatin remodeling and transcriptional regulation of a variety of genes [23].

PARP-1 is a highly conserved enzyme present in higher eukaryotes and *Archaea*[201] and is responsible for synthesis of the majority of poly(ADP-ribose) (PAR) within eukaryotic cells. Although typically highly expressed at the protein level (200,000-1 million copies per cell [103]), PARP-1 enzymatic activity is tightly regulated in the absence of genomic stress [202-205]. Under normal conditions, PARP-1 is found associated with histones, DNA and other chromatin associated factors [109, 206]. In response to DNA damage, PARP-1 enzymatic activity is stimulated 10-500 fold. This results in the poly ADP-ribosylation (PAR) of many target proteins including itself, core histones, linker histone H1 and transcription factors. The negative charge associated with PAR renders the target proteins unable to

associate with DNA and allows exposure to the cellular DNA repair machinery [109]. Interestingly, the PAR moiety itself is bound and hydrolyzed by many other enzymes including macro domain proteins and glycohydrolases [3,7,13,31].

While these processes are critical to genomic stability, the molecular mechanisms of how DNA damage induces PARP-1 activation are poorly understood. Due to the specificity of PARP-1 for non-B-DNA[207], we hypothesize that the location and orientation of PARP-1 when it is bound to damaged DNA is critical for subsequent activation of the PARP-1 enzymatic domain.

Human PARP-1 is a modular protein organized into at least 6 domains [47] (**Figure 4.1**). Its affinity for damaged DNA is regulated by two unique zinc-finger motifs ($CX_2CX_{28-30}HX_2C$) (residues 1-209; referred to as zf-PARP), which are sufficient to target the entire protein to damaged DNA [140, 208]. Recently, comparative analysis has defined the zf-PARP fold as an exceptional zinc-finger domain composed of a hydrophobic core with three conserved beta-sheets and two terminal alpha helices¹⁶. All zf-PARP family members contain two variable loop regions bridging $\beta 2-\beta 3$ and $\alpha 1-\alpha 2$ that may play a role in determining DNA structure specificity [207]. Proteins with the zf-PARP fold are evolutionary conserved in the eukaryotic lineage and are associated with many nuclear enzymes outside of the PARP protein family. None of the zf-PARP fingers are known to have sequence specificity, but instead seem to recognize many different DNA structures such as stem-loops, cruciforms, nicked DNA and 3'-extensions. The two PARP-1 zf-PARP domains are known to recognize DNA nicks, overhangs, blunt ends and other forms of damage [208-211]. Thus zf-PARP differ both structurally and in their DNA binding specificity from the abundant and well-studied family of sequence specific C2H2 zinc fingers [207]. However, little is known about the interactions of zf-PARP fingers with DNA [207].

Two recent publications have defined a third zinc domain in human PARP-1 [47, 48]. This zinc ribbon domain resides between residues 233-273 of the protein. Although the zinc ribbon domain does not bind DNA on its own, it is necessary for DNA stimulated activation of the full-length enzyme[47]. This has led to speculation that this domain modulates the N-terminal to C-terminal communication (possibly through a conformational change) that leads to DNA dependent activation of PARP-1 [47, 48].

Nevertheless, in the absence of multidomain PARP-1 structures, no direct evidence of a DNA dependent conformational change has been noted. Moreover, information on length requirements, affinity, stoichiometry and mode of interaction of the DNA sensing region in this abundant enzyme with various DNA damage models is lacking. Here, we characterize a highly purified fragment of human PARP-1 encompassing the entire DNA binding region, the zinc ribbon domain as well as the auto-modification domain (BRCT), with defined DNA damage models. The BRCT domain of PARP-1 is conserved with the BRCA-1 C-terminal region and is thought to function as a protein-protein interaction module. The structure and function all known human PARP-1 domains are reviewed in¹. The PARP-1 construct used in this manuscript was designed based on known domain structures Zn1, Zn2, Zn3 and BRCT (residues 1-486, from now on referred to as hparp486) (pdb IDs: 2dmj, 2cs2, 2jvn, 2cok) (Figure 1A). We find that hparp486 is a monodisperse monomer in solution on its own as well as in complex with blunt ended, nicked, or 3'-overhang DNA. Control experiments indicate that full length PARP-1 is activated to similarly high levels by the different DNA damage model substrates. Small angle X-ray scattering was used to determine the overall shape of hparp486 on its own and in complex with these DNA fragments. Distance distribution functions and *ab initio* particle reconstructions created from the SAXS data suggest a conformational change occurs in the zinc ribbon domain of the extended polypeptide chain upon DNA binding.

Materials and Methods

Cloning, expression and purification of hparp486:

DNA encoding the first 486 residues of human PARP-1 (codon optimized for E. coli) was cloned into the pET28a vector system (Novagen). hparp486 was expressed (via IPTG induction) in E. coli BL21 (DE3) (Novagen) for 6 hours at 25°C. Following cell lysis, the protein was bound to a Nickel-NTA fast flow column (Qiagen) and washed with 3 volumes of resuspension buffer prior to being eluted with 300

mM imidazole. hparp486 was further purified by S200 size exclusion and Hitrap SP cation exchange columns (GE Healthcare).

Cloning, expression and purification of Hparp209:

The DNA sequence encoding the first 209 residues of human PARP-1 (codon optimized for insect cells) were cloned into pAB-GST (AB-Vector) baculovirus transfer vectors and co-transfected with linearized Pro-fold C1 (AB-Vector) baculovirus DNA. The transfected virus was expanded and plaque assays were used to determine the isolates that best expressed hparp209. The resulting infectious viruses express hparp209 under the control of the very-late Polyhedron promoter. One liter of SF9 insect cells were grown to a density of 1×10^6 cells/ml in Ex-cell 420 serum free media (Sigma-Aldrich) and infected with an MOI of 3:1 for 72 hours at 28° C. At this point cell pellets were spun down and frozen at -20°C until further use. Upon thawing, the cell pellets were resuspended in lysis buffer (25 mM Tris HCl, 200 mM NaCl) and sonicated for 2 minutes. The suspension was spun at 10,000 g for 20 minutes and the supernatant was incubated with 10 mls of a Glutathione resin slurry (Qiagen) for 1 hour before washing with 50 mls of lysis buffer. The GST moiety was removed by incubating the protein with 100 units of PreScission Protease (GE Healthcare) overnight while rocking at 4°C. The protein was purified to homogeneity by HPLC chromatography over a Heparin HP ion exchange column. The protein bound at pH 7.0, 100mM NaCl and was eluted at pH 7.0 300mM NaCl (GE Healthcare). Size exclusion chromatography (S200; GE Healthcare) was used as a final purification step. The resulting protein preparation yields approximately 2.5 mg of hparp209 per liter of culture and the protein was stable for several weeks at a concentration of 1mg/ml when kept at 4°C.

Light Scattering Measurements:

2mg/ml samples of either hparp486, DNA or the hparp486-DNA complex in 75 mM Tris/HCl pH 8.0, 300 mM NaCl and 1 mM TCEP were loaded onto an ÄKTA purifier HPLC system. Samples were

characterized over a SHODEX-803 size exclusion column at a flow rate of 0.5 mls/minute prior to flowing into an online Dawn Heleos II (Wyatt Technologies) multiangle light scattering instrument followed by a REx refractive index detector (Wyatt Technologies). A differential index of diffraction (dn/dc) value of 0.185, 0.170, 0.180 were used to determine the concentration of hparp486, DNA and the hparp486-DNA complex respectively.

Analytical ultracentrifugation:

Samples were dialyzed extensively against 25 mM Tris pH 8.0, 200 mM NaCl and 1 mM TCEP prior to sedimentation. All experiments were performed in a Beckman XL-I using the absorbance optical system and a 4-hole, AN60-Ti rotor. Sedimentation velocity (SV) was performed in a 1.2 cm, 2-sector EPON centerpiece, while sedimentation equilibrium (EQ) was performed in a 1.2 cm, 6-sector centerpiece. For SV, 400 μ l samples of hparp486 were sedimented at 50,000 rpm for 4 hours at 20°C, with a radial step size of 0.002cm in the continuous scanning mode. For EQ, 100 μ l, samples were spun to equilibrium at 17,000, 24,000, 30,000 and 35,000 RPM.

DNA oligomer preparations:

The oligomer sequences for the 30Ext primer were 5'-ATCAGATAGCATCTGTGCGGCCGCTTAGGGTTAGGG-3' for the overhang strand and 5'-CCCTAAGCGGCCGCACAGATGCTATCTGAT-3' for the shorter strand. Other DNA constructs were generated by adding or deleting nucleotides from the 3' region for the sequences shown above. Oregon Green labeled 3'- and 5'- labeled and unlabeled sequences were ordered from IDT DNA. Various 3' overhangs, nicked and blunt ended double stranded DNAs were created by annealing equimolar amounts of oligomer. To anneal DNAs, the oligomer were mixed in an Eppendorf tube that was inserted in a 1 liter boiling water bath for 5 minutes, then allowed to cool to room temperature over night.

Mobility shift assays:

Protein-DNA complexes were assayed for homogeneity and stoichiometry by analyzing their mobility on 0.8-1.2% native agarose gels stained with Coomassie blue and ethidium bromide to visualize protein and DNA respectively. Low EEO agarose was obtained from Fisher. Gels were run in 25mM Tris pH. 8.0, 20 mM Glycine, at 60 volts for 60 minutes at 4°C.

Thermodynamic Measurements:

Affinity measurements of human parp486-DNA complexes were carried out in the Perkin Elmer Victor 3V plate reader. Corning 384 well non-binding polystyrene plates were used to minimize nonspecific binding of the protein complex to the chamber surfaces. In these experiments, human parp486 and DNA were resuspended into 75 mM Tris pH 8.0, 300 mM NaCl and 1 mM TCEP. 2 nM Oregon Green labeled DNA, 0.005-6μM PARP was added to each sample. The fluorophores were excited at 488 nm and the quenching of fluorescence at 530 nanometers was monitored. Triplicate samples were averaged and plotted with Kaleidagraph.

$$\text{Equation 1: } \text{Norm.F.C.} = F.C \max \left(\frac{P_t^{NH}}{P_t^{NH} + K_d} \right)$$

Data were fit to Equation 1, derived from reaction scheme 1 (see text) to the normalized fraction quenched (F.C.) which was observed as a function of P , where P_t is the total concentration of protein titrated, NH is the Hill coefficient, and K_d is the apparent dissociation constant. The NH was assumed to be one unless the data dictated otherwise.

Small Angle X-ray Scattering:

SAXS data for human parp486, DNA constructs and their complexes were measured at the SIBYLS beamline (12.3.1) at the ALS using a Mar CCD area detector (165 mm diameter) at room temperature. The intensity curves were measured at concentrations of human parp486 at 9mg/ml, 6mg/ml, and 3mg/ml; DNA samples at 6mg/ml and 3mg/ml; and human parp486-DNA complex samples equal to 6mg/ml and 3mg/ml. Data images were subjected to circular integration, normalization, and subtraction of sample and buffer image files. The R_g for each particle was approximated using PRIMUS (Kovnarev et al., 2003) to evaluate the Guinier equation and GNOM (Svergun, 1992) to evaluate the $P(r)$ function. The value of the maximum diameter of the particle, D_{max} , was determined empirically by examining the quality of the fit to the experimental data for a range of D_{max} values.

Ab initio particle reconstructions were built by the program DAMMIN [212]. For each final model, thirty independent DAMMIN reconstructions were aligned and averaged with the programs DAMAVER and SUPCOMB to reduce the likelihood that the models represent local minima [213, 214]. The most typical averaged model was filtered to an appropriate volume with the program DAMFILT.

In an independent approach, a model of human parp486 in solution was generated by rigid-body refinement from the high-resolution structures of the individual domains, obtained from the Protein Data Bank. The program SASREF [215] uses simulated annealing to find an optimal configuration of the domains by simultaneous fitting of the SAXS data to a model consisting of the coordinate files from Zn1, Zn2, Zn3 and BRCT domains (PDB entries 2dmj, 2cs2, 2jvn, 2cok, respectively). The coordinate files were initially truncated to minimize extended loops at either N or C terminal regions.

Results:

hparp486 is an elongated monomeric particle in solution

An N-terminally 6X His-tagged fusion construct of hparp486 was expressed in *E. coli* and purified to homogeneity. Purified hparp486 elutes in a single peak from a SHODEX-803 gel filtration column and SDS-page analysis reveals that the protein is ~95% pure (**Figure 4.1B**). Static light

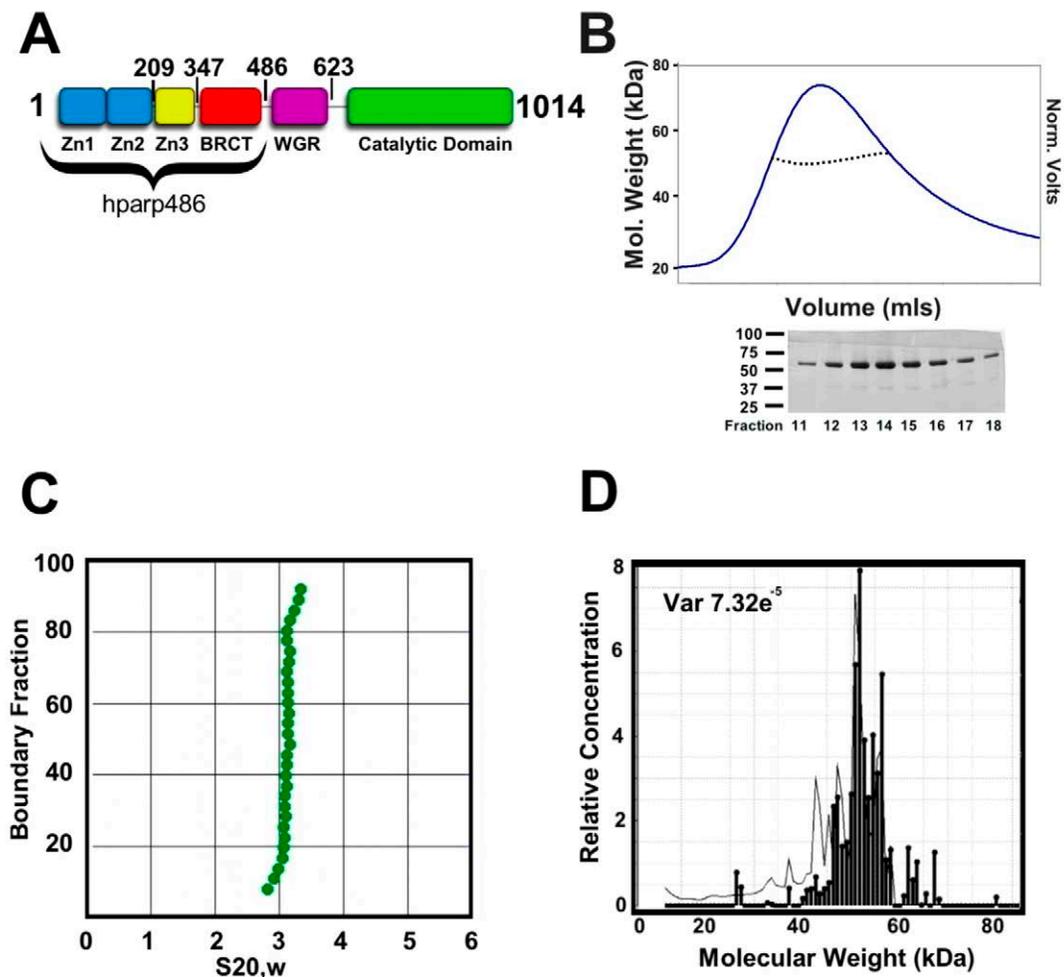


Figure 4.1. hparp486 is a monomer in the absence of DNA. **A)** hparp486 was designed to include the entire DNA binding region of the human PARP-1 protein. The first 486 residues encompass the three N-terminal Zn domains and the BRCT domain. The illustration indicates the C-terminal residue of each domain. **B) Top:** Size exclusion chromatography and inline static light scattering of hparp486. hparp486 elutes in a single peak from a KW803 size exclusion column. The static light scattering average molecular weight (~57kDa) for the central portion of the peak is illustrated by the black dotted line within the peak. **Bottom:** SDS-PAGE analysis of fractions from SEC purified hparp486. **C)** The sedimentation velocity profile of purified are presented as G(s) plots of the integral of $S_{20,w}$ against boundary fraction (%). van Holde - Weischet analysis was used to determine the f/f_0 value of 1.8. **D)** Sedimentation equilibrium ultracentrifugation resolved that over 95% of the purified protein was monomeric at a concentration of 40 μ M.

scattering measurements determined that the purified protein is a monomer with a molar mass of ~57 kDa at concentrations exceeding 2 mg/ml (**Figure 4.1B**). Sedimentation velocity experiments determined that hparp486 sediments as a single species of elongated character with an anhydrous frictional ratio (f/f_0) = 1.8 (**Figure 4.1C**). Further sedimentation equilibrium experiments demonstrate that hparp486 has a molecular mass of 55 – 60 kDa (**Figure 4.1D**). These results are in excellent agreement with the calculated mass of a hparp486 monomer, 56.480kDa. Small angle X-ray scattering (SAXS) is an effective method for determining the low-resolution ($>10\text{\AA}$) structures of proteins, nucleic acids and their complexes in solution at reasonably low concentrations. After an overall SAXS envelope is determined, hybrid methods can be applied to fit domains of known structure within the SAXS envelope [216]. SAXS data were collected on hparp486 at 3, 6 and 9 mg/ml and the scattering curves were overlaid to determine any concentration-dependent effects. To determine the relative amount of folded domains and any aggregation within the hparp486 sample, the data were plotted in a Kratky plot and analyzed by Guinier approximation respectively (**Supplemental Figure 4.1**). From this preliminary analysis we determined that the data were deemed of sufficient quality to use the scattering profiles up to $q_{\text{max}}=0.205\text{\AA}^{-1}$ (where q is the scattering vector). For hparp486, the radius of gyration varies with increasing concentration in the range $R_g \sim 46-48\text{\AA}$ (**Table 4.1**). Further analysis of the hparp486 scattering data determined that the particles in solution do follow Porod's law, reviewed in [214]. Thus their volume can be calculated from their scattering profile. We established that hparp486 occupies $\sim 103,770 \text{\AA}^3$ and has a molecular weight of $\sim 54 \text{ kDa}$ (**Supplemental Figure 4.1**). This is in good agreement with values that were determined for hparp486 by AUC ($\sim 60 \text{ kDa}$) and classical light scattering at lower concentrations ($\sim 57 \text{ kDa}$; **figure 4.1**). Thus, hparp486 exists as a monodisperse monomer over a wide range of concentrations (3-9mg/ml).

Additional information can be extracted from well-behaved SAXS data. The GNOM program [217] calculates pair distance distribution functions (denoted as $P(r)$) from scattering data using an indirect Fourier transformation procedure. For hparp486 we observe two maxima in the $P(r)$ function

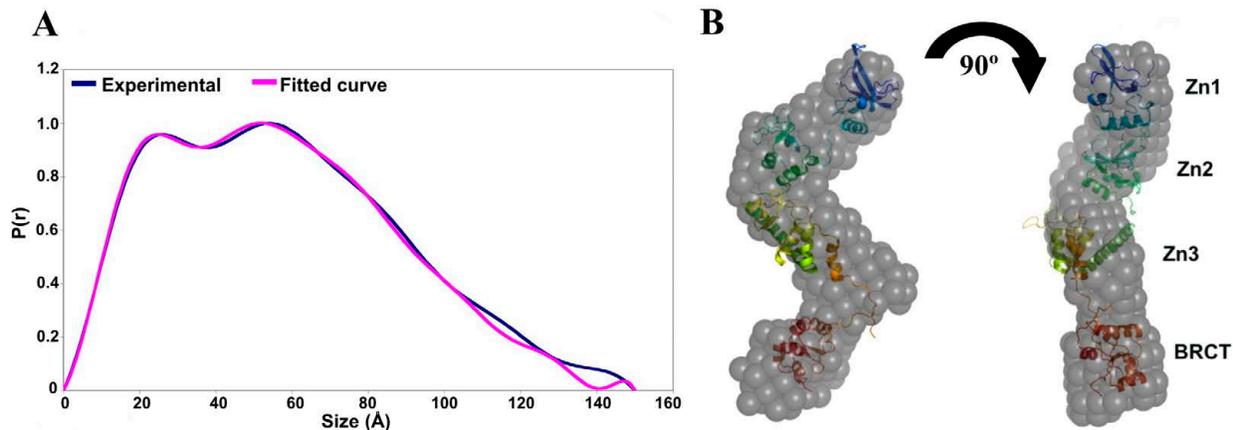


Figure 4.2. The low resolution structure of hparp486 reveals a molecule with conformational flexibility. **A)** Superposed experimental and modeled distance distribution functions of hparp486. In blue, the SAXS distance distribution function $P(r)$ for hparp486 is an asymmetric curve with two discrete maxima, representing two ordered regions connected by a flexible linker. The relatively shallow descent towards a maximum size of 150Å is reflective of an elongated, flexible molecule. Illustrated in violet, is the $P(r)$ function resulting from the SASREF tertiary model shown in (B). **B)** Superposed particle reconstructions and tertiary structure models of hparp486. The DAMMIN particle reconstruction (grey) represents the filtered average of 30 models individually computed from a SAXS scattering curve and fits the SAXS scattering curve with a χ^2 of 1.10. The SASREF tertiary structure model was computed by global rigid body modeling of the known domain structures of Zn1 (PDB id 2dmj, blue), Zn2 (PDB id 2cs2, green), Zn3 (PDB id 2jvn orange) and BRCT (PDB id 2cok, red) with steric restraints against the solution scattering data; this model has a χ^2 of 1.21. In the orientation shown, the two models have a correlation coefficient of 0.56, while a 180° rotation of the polypeptide resulting in a switch in position of Zn1 and BRCT domains has a correlation coefficient of 0.54. The SASREF model illustrates the extended nature of the polypeptide and minimal inter-domain contacts.

Table 4.1. Overall dimensions hparp486, DNA and hparp486-DNA complexes. Molar masses were obtained by SEC-MALS, radii of gyration (R_g) and maximum dimension (D_{max}) were obtained from SAXS. The calculated molar masses of each molecule and a 1:1 complex of hparp486-DNA are in the second column. The SEC-MALS determined mass listed in the third column. The radius of gyration (R_g) for each molecule at the lowest concentration tested (3 mg/ml) while the highest concentration (9 mg/ml for hparp486 and 6 mg/ml for all other samples) are listed in the fourth and fifth columns respectively. The last column contains the maximum dimension for each molecule as determined by SAXS. Variation of D_{max} was negligible for these molecules within this concentration range.

| Molecule | Calculated Mol. Mass | SEC-MALS Mol. Mass | SAXS R_g (Å) [3mg/ml] | SAXS R_g (Å) [6 or 9 mg/ml] | SAXS D_{max}(Å) |
|-------------------------|-----------------------------|---------------------------|---|---|-------------------------------------|
| hparp486 | 56.0 | 54.4±0.1% | 46.3 | 48.3 | 150 |
| Hparp209 | 23.0 | 22.5±3% | 29.0 | 29.5 | 115 |
| 30Nick | 18.4 | 22.8±5% | 26.9 | 28.6 | 90 |
| 30Blunt | 18.4 | 22.0±2% | 28.2 | 28.5 | 88 |
| 30Ext | 20.2 | 21.2±0.1% | 30.2 | 29.8 | 100 |
| hparp486-30Nick | 74.4 | 83.5±6% | 63.5 | 64.2 | 195 |
| hparp486-30Blunt | 74.4 | 70.0±6% | 76.5 | 78.7 | 240 |
| hparp486-30Ext | 76.3 | 71.3±1% | 63.5 | 63.0 | 195 |
| Hparp209-30Blunt | 43.1 | --- | 55.2 | 51.6 | 180 |

(**Figure 4.2A**). This is characteristic of two distances appearing with a higher probability, and indicates a macromolecule composed of at least two ordered domains connected by a flexible linker. Additionally, the $P(r)$ function shows a shallow descent between 125 and 150Å, which is the signature of an elongated, flexible particle.

Two models of hparp486 were independently calculated with the programs DAMMIN and SASREF [212, 215]. The DAMMIN program was used to calculate 30 ab initio models that fit the SAXS scattering profile of hparp486 with a $\chi^2=1.0-1.1$. A superposition of 5 randomly chosen models is shown in supplemental **figure 4.2**. The filtered, averaged envelope of the 27 most similar models has a maximum dimension (D_{\max}) of ~148Å and forms an overall S-shaped structure (**Figure 4.2B**). Most residues (432 of 486, or 90%) within the N-terminal half of PARP-1 are represented in one of four domains previously deposited in the Protein Data Bank (PDB IDs: 2dmj, 2cs2, 2jvn, 2cok). The SASREF program performs tertiary structure modeling by reconstructing all or parts of the SAXS model from domains of known structure. The PARP-1 domains were trimmed of extended linkers and subjected to N-C terminus distance restrictions that limited overlapping regions and minimized steric restraints during the modeling process. The domain definitions and optimized restrictions are summarized in the Materials and Methods section. The SASREF model has a D_{\max} of ~140Å and fits the hparp486 $P(r)$ function with a $\chi^2=1.21$ (**Figure 4.2A**). The DAMMIN and SASREF models superpose well (**Figure 4.2B**), and it is clear that they both represent an overall S-shaped elongated molecule.

The hparp486 interacts with damaged DNA as a monomer

Previous biochemical studies have monitored the interaction of truncated PARP-1 constructs with 55 base pair DNA fragments[210, 211]. We initially assayed the association of hparp486 to a 53 base pair blunt ended DNA by electrophoretic mobility shift analysis (EMSA) and found that the protein could

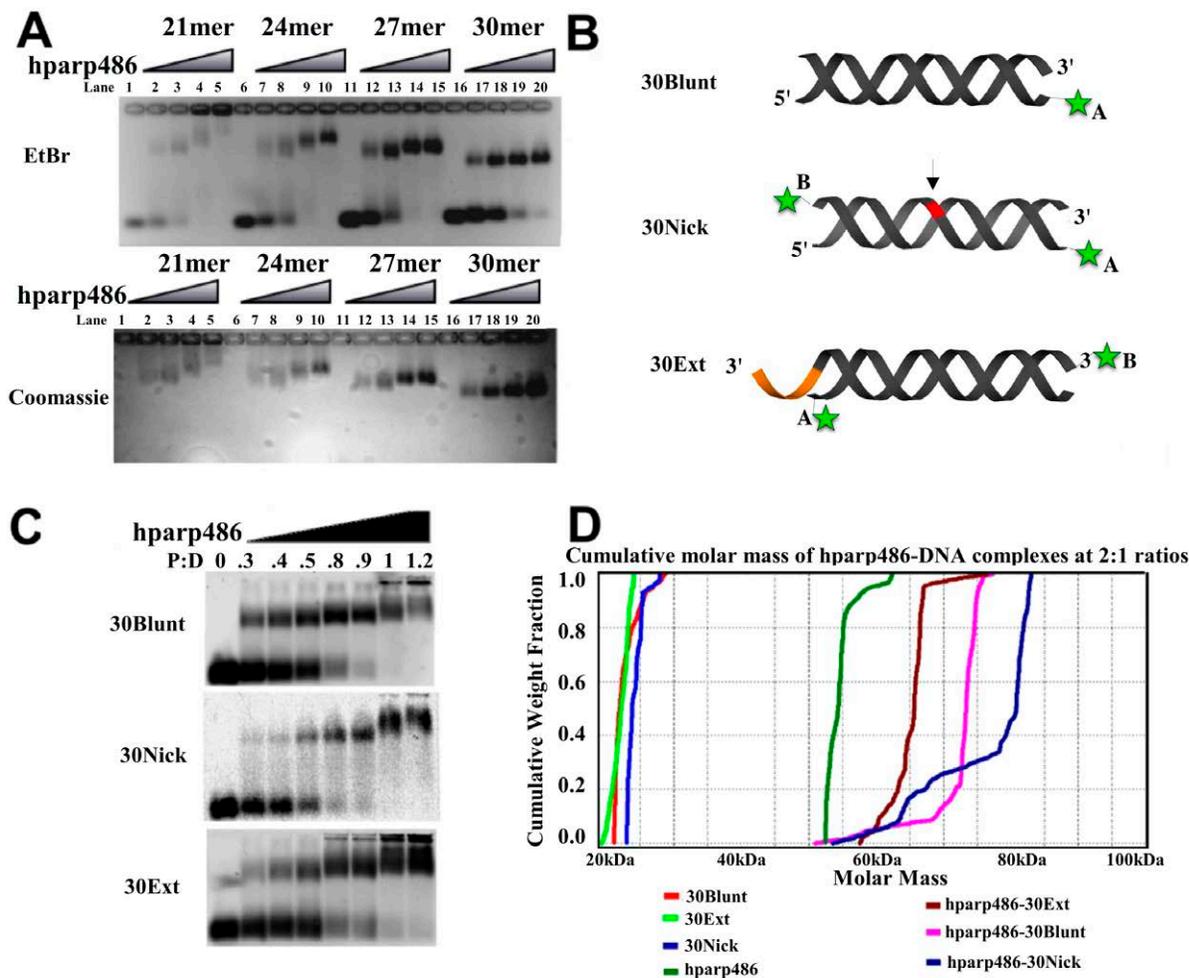


Figure 4.3. hparp486 binds as a monomer to damaged DNA. **A)** EtBr and Coomassie stained 1% agarose gel EMSA assay of hparp486-DNA complexes. $1\mu\text{M}$ DNA was added to each lane. In this assay, all DNA constructs contain a 6-nucleotide 3' extension, but vary in the length of the double stranded region (21-30 base pairs). hparp486 forms homogenous complexes as the DNA approaches 27-30 base pairs in length. **B)** A schematic representation of the three DNA damage models tested and the location of the fluorophores (green asterisk) used in this publication. Varying the fluorophore position (either in position A or B) allowed us to test the effect of its position on hparp486 association. The arrow and red region highlight the area of the single stranded break in the 30Nick DNA, while the orange area highlights the 6 nucleotide 3' extension in 30Ext. **C)** EMSA of hparp486 association with $2\mu\text{M}$ fluorescently labeled 30Blunt, 30Nick and 30Ext DNA. hparp486 efficiently shifts the mobility of over 90% of each DNA at a 1:1 ratio, eliminating the possibility of cooperative dimerization under these conditions. **D)** Cumulative molar mass of DNA, hparp486 and hparp486-DNA complexes as determined by SEC-MALS. 2:1 molar ratios of hparp486:DNA were incubated for 30 minutes prior to subjecting the mixtures to SEC-MALS. The distribution of masses present in the major peaks eluting from the column for DNA, hparp486 or each hparp486-DNA complex is plotted as a fraction of cumulative molecular weight. hparp486 does not stably dimerize on any DNA fragment. Supplemental Figure 4 contains the elution profiles for each solution.

form either 1:1 or 2:1 complexes with this DNA at hparp486:DNA ratios between 0.5:1 and 2:1 (**Supplemental Figure 4.3**). Therefore, we first set out to determine the base pair length requirements needed to promote stable binding in order to investigate whether hparp486 binds DNA as a monomer or dimer. We used an agarose gel-based mobility shift assay in which we visualize the homogeneity of various hparp486-DNA complexes. Shown in **Figure 4.3A** is the effect of double-stranded DNA length on the homogeneity of complexes formed between hparp486 and DNA containing 6-nucleotide 3'-overhangs. As the double-stranded DNA length approached 27-30 base pairs, the complex migrates as a sharp single band when visualized by both EtBr and coomassie stains as opposed to a weaker, more diffuse shifting pattern observed with shorter DNA. We concluded from the mobility-shift assay that a 30 base pair double helical region is sufficient to stably bind hparp486 when a 6-base 3'-overhang is present. In order to determine how hparp486 interacted with other forms of DNA damage, we continued the remainder of our experiments to compare DNA of the same base pair length, but with varying ends (blunt ends, and nicked DNA).

To determine the stoichiometry of the various hparp486-DNA complexes, we used 2 μ M fluorescently labeled DNA in an EMSA to quantify the ratio at which all DNA formed a complex with hparp486. DNA molecules with a single label at either location A or B were designed for the purpose of testing the stoichiometry and affinity of hparp486 association (**Figure 3B**). We determined by EMSA that when using 30 base pair DNA, hparp486 forms homogeneous 1:1 complexes irrespective of nicks, extensions or blunt ends (30Ext, 30Nick, and 30Blunt; Figure 3C). The addition of hparp486 past a 1:1 protein to DNA molar ratio resulted in the retention of both protein and DNA within the wells irrespective of the length of time the gel was run. This may represent either aggregation or a low affinity, possibly anti-cooperative dimerization of hparp486 on these DNA fragments. To overcome the limitation of the mobility shift assay, we confirmed the stoichiometry of hparp486 on DNA in solution by Size Exclusion Chromatography in conjunction with Multi Angle Light Scattering (SEC-MALS). A 2:1 mixture of hparp486 to DNA was loaded on a SEC column that eluted directly into a MALS instrument. An average

Table 4.2. SASREF domain definitions and distance restrictions. The solution structures for the human PARP-1 domains Zn1, Zn2, Zn3 and BRCT (PDB entries 2dmj, 2cs2, 2jvn, 2cok, respectively) were truncated of N- and C-terminal flexible regions prior to tertiary model reconstruction by the program SASREF. The residues included and C-terminal distance restrictions used in the tertiary model are listed for each domain.

| Domain | Modeled Residues | C-term distance restriction |
|---------------|-------------------------|------------------------------------|
| Zn1 | 1-193 | 20Å |
| Zn2 | 103-207 | 7Å |
| Zn3 | 225-359 | 30Å |
| BRCT | 380-486 | ---- |

molecular weight for macromolecules within each peak eluting from the SEC column was calculated from the MALS data. hparp486 has a molar mass of ~56 kDa while each DNA is ~20 kDa, thus 1:1 complexes should be ~76 kDa, while a 2:1 complex would be ~133 kDa. SEC-MALS determined the molar masses for the hparp486 in complex with 30Ext, 30Blunt and 30Nick were 71.3, 74.4 and 84.5 kDa respectively; no evidence for a 2:1 complex was detected (**Table 4.1**). In line with this result, the cumulative weight fraction and molar mass vs. volume profiles determined from the SEC-MALS experiments both illustrate the presence of a significant amount of free hparp486 in solution (**Figure 4.3D, Supplemental Figure 4.4**).

We next determined the affinity of hparp486 for the 30 base pair DNA damage models by monitoring the quenching of the fluorescence of labeled DNA molecules in solution. Using Equation 1 (materials and methods) we fit the single-step scheme $P + D \xrightleftharpoons{K_1} PD$ (scheme 1, P= hparp486, D=DNA $K_1=K_d$) to the experimental data obtained for each protein-DNA complex. We determined that under uniform conditions, hparp486 associates with 30Ext \approx 30Nick (K_d of ~100nM) > 30Blunt (K_d of ~300nM), when they all contain the same 30bp ‘DNA core’ (**Figure 4.4A, Table 4.3**). To minimize the possibility of the fluorophore interfering with protein – DNA interactions, we repeated the experiment with 30Ext with the fluorophore at multiple locations without a significant change in affinity (**Table 4.3**). However, when the 30Nick DNA was labeled at position A, interacting with hparp486 did not result in sufficient quenching of fluorescence (8%) to determine the K_d (**Table 4.3**). Thus hparp486 does not change the environment of the fluorophore at the blunt end when it associates with the region surrounding the nick. We determined that hparp486 associates with each DNA with a Hill coefficient of 0.92-0.99, reflecting the lack of a cooperative association between the protein and any of these DNA damage models (**Table 4.3, Supplemental Figure 4.5**), and further supporting the notion that hparp486 binds as a monomer.

hPARP-1 is activated by DNA containing nicks, 3'-extensions and blunt ends

The hPARP-1 enzyme is stimulated 10-500 fold upon binding damaged DNA [218]. We used a

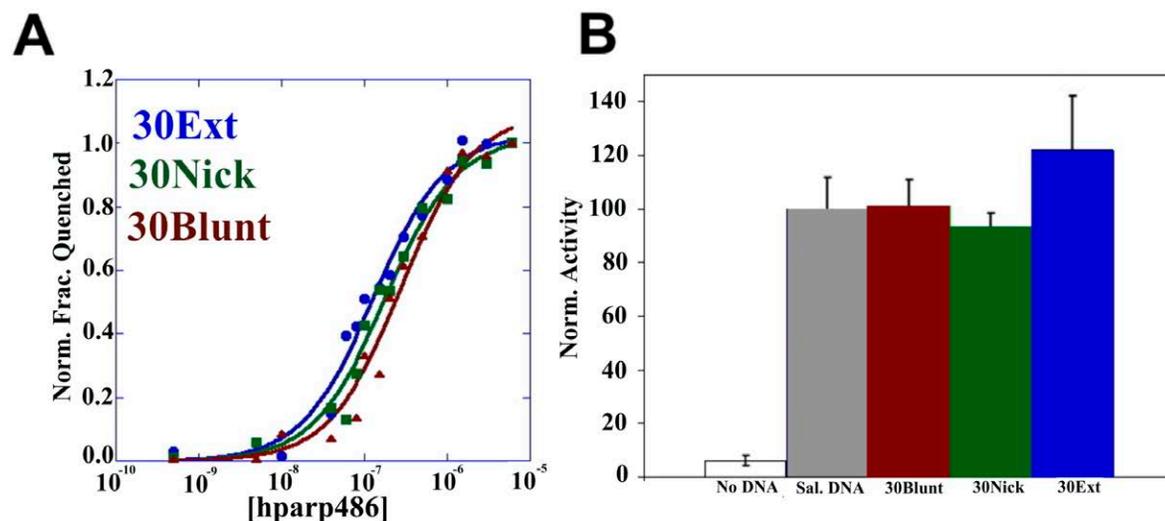


Figure 4.4. hparp486 has similar affinities for different forms of DNA damage; full-length human PARP-1 is efficiently activated by DNA containing blunt ends, nicks or a 3'-extension. **A)** The affinity of hparp486 for fluorescently labeled 30 base pair DNA: 30Blunt (red), 30Nick (green), 30Ext (red) was determined by measuring the change in fluorescence of 2nM labeled DNA as a function of hparp486 concentration (0.5-6000nM) in solution. The affinity of hparp486 each DNA is shown in **Table 3** and is in the range of ~70-300nM. The Hill coefficients calculated for each fluorescent titration were in the range of 0.92-0.99, demonstrating the lack of positive cooperativity in the individual interactions (**Table 4.3**). **B)** The ability of each DNA to activate full-length human PARP-1 was tested with a commercial assay that monitors the incorporation of biotinylated ADP-ribose onto immobilized histones. The activation induced by 1 μ M of 30Blunt, 30Nick or 30Ext DNA was normalized to the activity stimulated by sheared salmon sperm DNA. Regardless of the DNA damage model tested, PARP-1 was ~40 fold more active than in the absence of DNA.

Table 4.3. The affinity of hparp486 for blunt, nicked and 3'-extension DNA. The percent fluorescence quenched, affinity (Kd) and Hill coefficient determined from the interaction of hparp486 with fluorescently labeled blunt ended, nicked or 3'-extension DNA are listed. The DNA constructs were either labeled in Position A or B, as illustrated in Figure 3B. Because of its pseudo symmetry, the 30Blunt construct was only labeled in one location.

| DNA Type | Fluorescence Quenched | Affinity (Kd) | Hill Coeff. |
|------------------------|-----------------------|--|-------------|
| 30Blunt | 33% | $2.82 \times 10^{-7} \pm 4.0 \times 10^{-8}$ | 0.9206 |
| 30Nick (pos. B) | 44% | $1.46 \times 10^{-7} \pm 2.0 \times 10^{-8}$ | 0.9670 |
| 30Nick (pos. A) | 8% | ---- | ---- |
| 30Ext (pos. B) | 35% | $1.30 \times 10^{-7} \pm 1.3 \times 10^{-8}$ | 0.9904 |
| 30Ext (pos. A) | 19% | $6.80 \times 10^{-8} \pm 6.0 \times 10^{-9}$ | 0.9802 |

chemiluminescent hPARP-1 assay kit (Trevigen Inc.) to measure hPARP-1 activity in the presence of the various 30 base pair DNA fragments tested above. This assay measures the incorporation of biotinylated Poly (ADP-ribose) onto histone proteins in a 96-well plate format. 30Ext, 30Nick and 30Blunt DNA were each able to stimulate hPARP-1 ~40 fold over the basal activity in the absence of DNA and to the same level as sheared salmon sperm DNA (**Figure 4.4B**).

DNA binding induces conformational changes in hparp486

We used SAXS to obtain structural information on the various hparp486-DNA complexes. As a control, we measured the scattering curves for 30Ext, 30Blunt and 30Nick DNA individually at 3 and 6 mg/ml. While small concentration-dependent differences in radius of gyration (R_g) were noticed, superposition of the SAXS scattering profiles and Guinier analysis determined the data are of good quality (**Supplemental Figure 4.5**). Charge repulsion energies are known to affect DNA monodispersity in solution, and this may account for the observed concentration dependent effects [219]. The analysis of the $P(r)$ function illustrates that all DNA molecules conform to a double helix in either the A or B DNA form in solution, within the resolution of the technique (88-100Å in length). A comparison of the $P(r)$ functions for all datasets obtained for DNA is shown in **Supplemental Figure 4.6**. The DAMMIN particle reconstructions of each DNA molecule are illustrated in **Supplemental Figure 4.7**.

Despite similarities in binding affinities, it is possible that the various DNA damage models bind differently to hparp486. We collected SAXS data on hparp486 in complex with 30Nick, 30Ext or 30Blunt DNA. After initial protein and DNA concentrations were determined, each complex was further titrated to homogeneity as determined by mobility shift assays. SAXS data for hparp486 in complex with various DNA fragments were initially analyzed the same way as was done for data from hparp486 or DNA fragments alone: The dependence of the nature of scattering on concentration was studied by comparing scattering profiles (**Supplemental Figure 4.8A**), the scattering behavior at low angle was studied by Guinier analysis (**Supplemental Figure 4.8C**), the protein folding was checked by

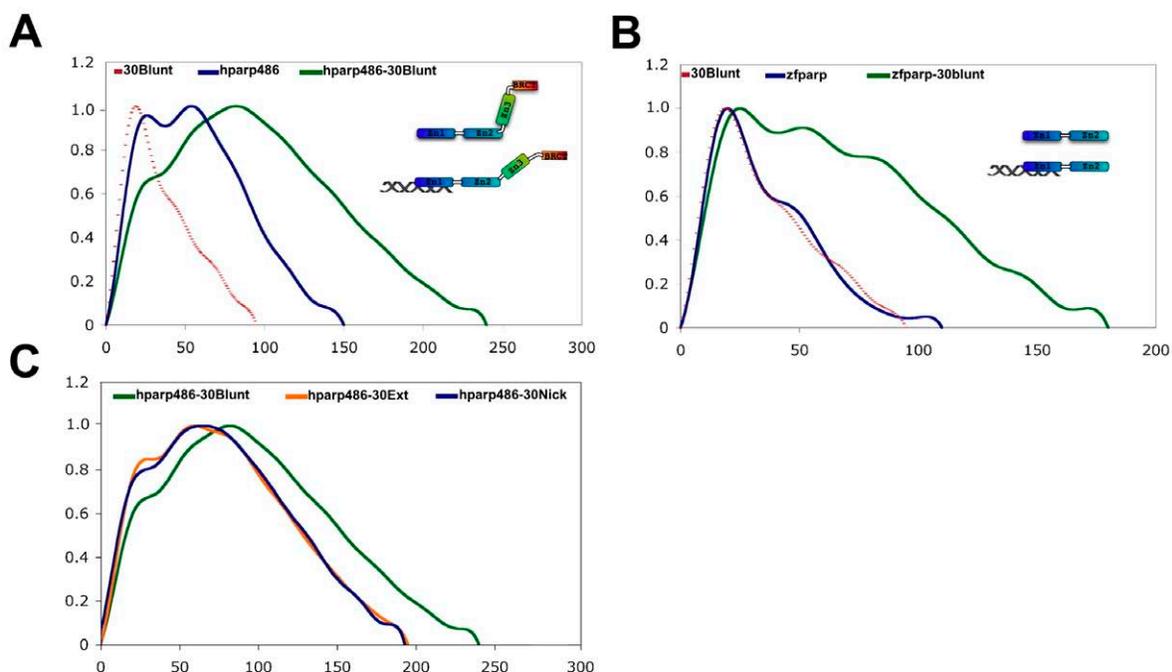


Figure 4.5. The distance distribution functions of hparp486-DNA complexes portray a conformational change within hparp486 upon binding DNA. **A)** The $P(r)$ functions of 30Blunt (dotted line), hparp486 (blue) and a 1:1 complex of hparp486-30Blunt complex (green). Upon binding 30Blunt, the profile of hparp486 in complex changes significantly when compared to the normalized profile of hparp486 in the absence of DNA. In particular, interacting with DNA changes the $P(r)$ function from two discrete maximum of similar height to a single dominant maximum of 240\AA , reflecting an increase in D_{max} of 90\AA over hparp486. We interpret this to represent a migration from a conformation containing two ordered regions connected by an internal flexible linker to an elongated, dynamic complex that does not have an internal highly flexible region. **B)** The $P(r)$ function of zf-parp (blue), 30Blunt (dotted line) and a 1:1 complex of zf-parp-30Blunt. In contrast to hparp486, the $P(r)$ function of zf-parp is characterized by a distinct single maxima, which reflects the lack of a high degree of internal flexibility. Upon binding 30Blunt DNA, no evidence of a conformational change is noted and with a maximum dimension of 180\AA , the complex reflects an increase in size of only 65\AA . **C)** The SAXS scattering profiles of solutions containing 1:1 ratios of hparp486 in complex with 30Blunt, 30Nick or 30Ext DNA were analyzed to determine the general shape and size of the complexes (green, blue, orange respectively). The $P(r)$ functions of the complexes illustrate that hparp486 forms complexes of similar maximum size with 30Nick and 30Ext DNA ($\sim 200\text{\AA}$), but forms a much longer complex with 30Blunt DNA ($\sim 240\text{\AA}$ maximum length).

representing the data in a Kratky plot (**Supplemental Figure 8B**) and the pair distance distribution functions were determined by indirect Fourier Transform (**Figure 4.5**). The results show that there is minimal concentration dependence for the SAXS data; the Kratky plot is typical for partially flexible molecules.

The SAXS scattering curves for each hparp486-DNA complex are depicted in **Supplemental Figure 4.8**. The $P(r)$ functions hparp486, 30Blunt and the hparp486-30Blunt complex allowed us to qualitatively compare the overall shape of particles in solution and quantitatively compare R_g and D_{max} values with the relatively rigid DNA fragment (**Figure 4.5**). A comparison of the $P(r)$ functions indicates a significant change in the particle nature upon binding of hparp486 to 30Blunt DNA. hparp486 alone consists of at least two folded domains (of ~ 40 and 60 \AA in size) connected by a flexible linker. On the other hand, the $P(r)$ function of hparp486-30Blunt DNA is distinct in almost all of its characteristics when compared to hparp486 on its own (**Figure 4.5A**). This profile has a sharp maximum at $\sim 100 \text{ \AA}$ and a D_{max} of 240 \AA (an increase of 90 \AA over hparp486 on its own). In contrast to the double maxima of the hparp486, the $P(r)$ function of the complex contains a single maximum, which could be reflective of a structural change within the protein that occurs upon binding DNA.

To identify if the changes in the $P(r)$ function were reflective of a conformational change-taking place in the Zn3-BRCT region of the protein, we repeated the experiment using the zinc finger region of PARP-1 (residues 1-209, referred to as hparp209). We found hparp209 behaved as a homogenous monomer by SEC-MALS (**Supplementary Figure 9A**), and bound 30blunt DNA as a monomer by mobility shift assay (data not shown). The raw SAXS scattering profile and Kratky plot of hparp209 are shown in **Supplementary Figures 9B-C**. The $P(r)$ function of hparp209 contains a single discreet maximum and a smaller shoulder peak at $\sim 50 \text{ \AA}$ in size (**Figure 4.5B**). Particle reconstructions of hparp209 clearly depict a flexible terminal region of the protein, but otherwise superimpose well with each other (**supplementary figure 4.9D**). The hparp209-30Blunt complex maintains the same overall profile as hparp209 on its own but only increases in size by 65 \AA to a D_{max} of 180 \AA (**Figure 4.5B**). Thus,

in the absence of Zn3-BRCT, hparp209 retains the ability to interact with DNA, but does not retain the characteristic change in shape that we identified in hparp486.

Inspection of the $P(r)$ function allows us to compare the complex formed between hparp486, 30Ext, 30Nick DNA and 30Blunt (**Figure 4.5C**). We determined a similar D_{\max} for the 30Nick and 30Blunt complexes, but note some differences in the profile of the $P(r)$ function for each complex. We found that both particles have a D_{\max} of $\sim 190\text{-}210\text{\AA}$ and exhibit an overall elongated shape. Similarly, both complexes have a broad peak in the $P(r)$ plot with a maximum at $\sim 80\text{-}100\text{\AA}$ in size. At approximately $190\text{-}210\text{\AA}$ in D_{\max} , these complexes are $\sim 20\%$ shorter than the sum of D_{\max} obtained for hparp486 and DNA individually and the hparp486-30Blunt complex.

Discussion

The specific molecular mechanisms that lead to PARP-1 enzymatic activation in the presence of DNA damage have yet to be identified. Our understanding of the enzymatic complexity of PARP-1 relies not only on the structure and function of individual domains, but requires knowledge of the inter-domain communications that regulate the enzyme. Here we propose that PARP-1 zf-parp domains communicate the presence of DNA-damage by inducing a conformational change that leads to activation of the PARP-1. Additionally, we have established many basic features of DNA recognition by PARP-1 that were previously unknown, for example length requirements, affinity, stoichiometry, and the mode of interaction.

Notably, our SAXS data have provided the first low-resolution structure of the N-terminal half of PARP-1 (**Figure 4.2**). The most striking feature of this model is interior structural flexibility, (likely in the form of hinges connecting two structured domains or sets of domains), illustrated by the double maxima in the hparp486 $P(r)$ function. The region of interior flexibility is $\sim 55\text{\AA}$ from either the C or N-terminus of hparp486. After comparing the particle reconstruction, tertiary model and the dimensions of the individual PARP-1 domains, we propose two models for the sequences, which might constitute the

hinge region. The first would be a region immediately following the second zinc finger domain (residues ~208-220). The distance across the zinc-binding face of the first two zinc finger domains is ~25Å, and when including a small linker region between the domains the total distance between the N-terminus of zinc finger 1 to the C-terminus of zinc finger 2 is ~50-60Å. An alternative model places the hinge region between residues 370-385. This region is ~55Å away from the BRCT domain C-terminus in the hparp486 SASREF model. Interestingly, either model supports the notion that the zinc ribbon domain functions to relay the DNA damage signal from the zf-PARP region to the enzymatic domain. Moreover, both models are supported by the P(r) function of hparp209, which does not contain this interior flexible region.

We found that PARP-1 has similar affinities for the three DNA damage forms tested. This may be a result of PARP-1 having evolved as a general sensor of DNA ends and not as a specialized ‘nick’ sensor. The ~3 fold reduced affinity of PARP-1 for blunt ended DNA may be due to the increased rigidity of this DNA fragment, as two zf-PARP containing enzymes have been shown favor large DNA distortions [61, 220].

The SAXS data of the PARP-1-DNA complexes depict an overall elongated, asymmetric structure. In line with previous results[221], the D_{\max} of the PARP-1-DNA complexes suggest significant contacts are made between the protein and double stranded DNA in the three models that were tested. The asymmetric form of binding with which PARP-1 associates to the region surrounding the DNA damage is reminiscent of the DNA damage recognition of HU-like proteins within bacteria [222, 223]. Recently, HU-like proteins have been shown to associate with cruciforms, nicks and other high affinity DNA targets of PARP-1 [222, 223]. This asymmetric association to regions of DNA damage may represent a conserved mechanism that exposes the lesion while protecting the surrounding undamaged region.

Two recent structural studies of the PARP-1 zinc-ribbon domain have revealed that this domain is required for DNA-mediated activation of PARP-1 by either catalyzing dimerization of the full-length protein or inducing a conformational change[47, 48]. In light of our SAXS data and in the absence of any

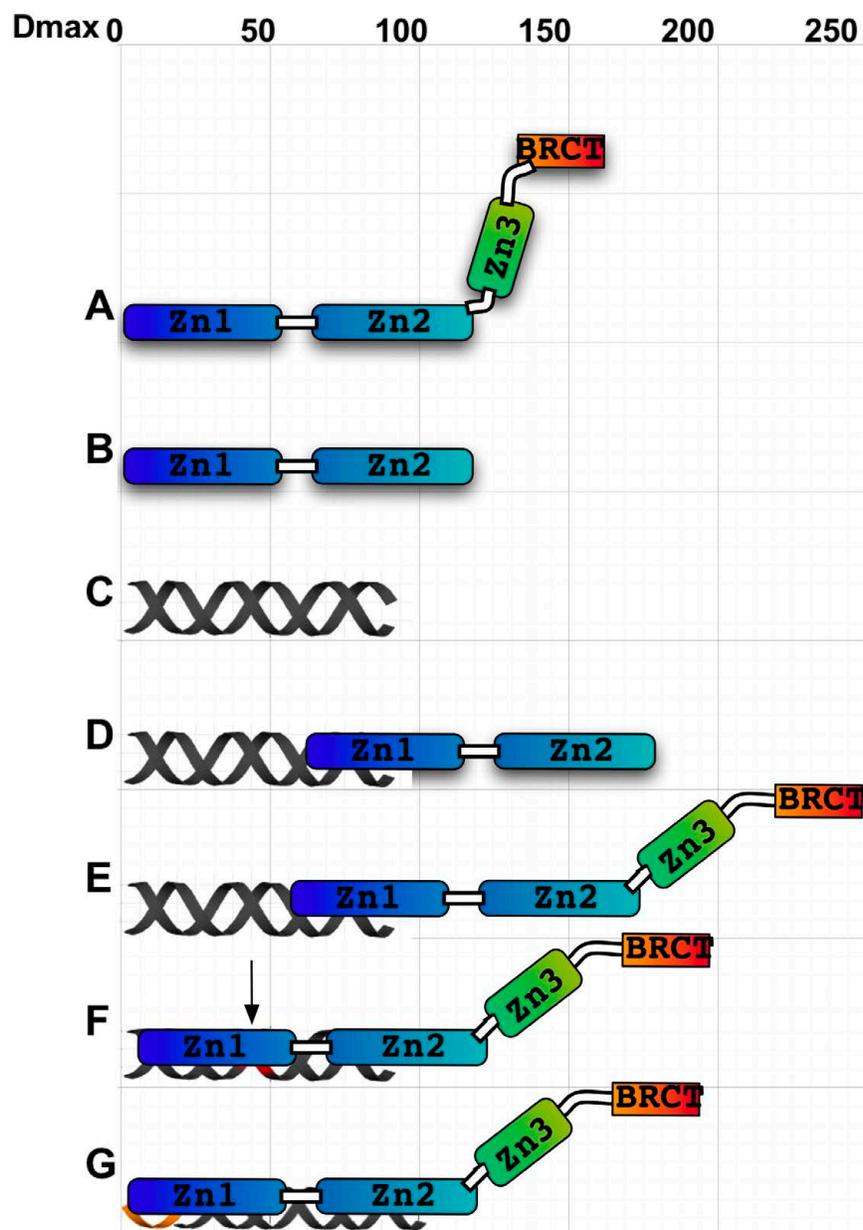


Figure 6. A model for the interactions of hparp486 with damaged DNA. The overall shape and D_{\max} of hparp486 (A), zf-parp (B), 30Blunt DNA (C) and their complexes (D and E) have been experimentally determined by SAXS. The D_{\max} of 180Å (a 65Å increase) and asymmetric shape of the zf-parp-30Blunt complex is consistent with a single parp zinc finger interacting the blunt end (D). The hparp486-30Blunt complex is also elongated (E), but increases in size by 90Å (to 240Å). To account for both the change in profile of the $P(r)$ function (Fig. 5 A, C) and greater D_{\max} , we model this with a conformational change that occurs in the Zn3-BRCT region of the protein. In contrast, when interacting with nicked DNA (F) or a 3' extension region (G), hparp486 forms a more compact complex but a conformational change still needs to be taken into account to reach a D_{\max} of ~200Å. We propose that ultimately the conformational change that occurs during binding of DNA leads to activation of the PARP-1 enzymatic domain.

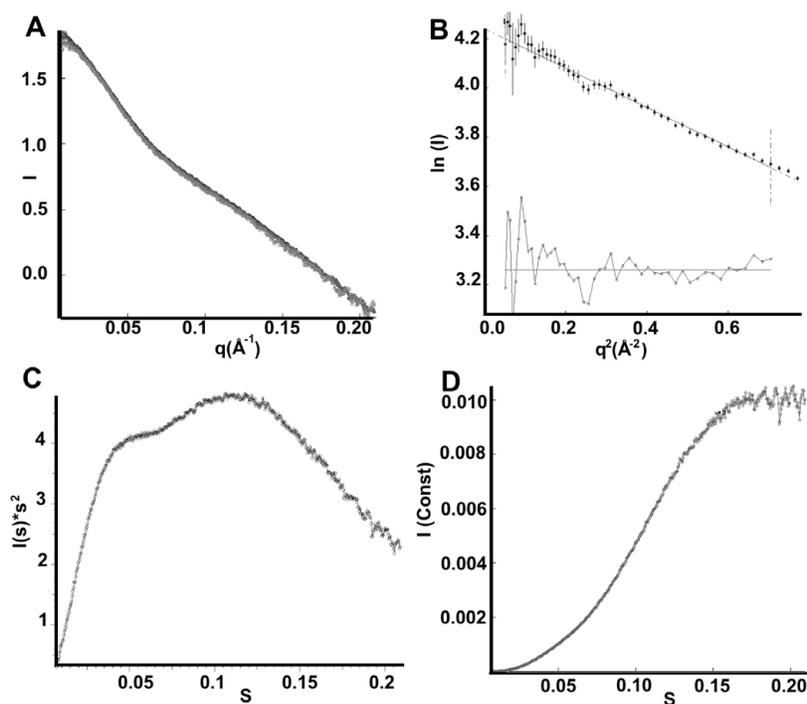
data suggestive of DNA-mediated dimerization, we support the notion that a DNA-mediated conformational change in the region surrounding zinc-ribbon domain leads to PARP-1 activation. We propose that the conformational change that occurs upon DNA association is communicated across the hinge region, and ultimately leads to activation of the PARP-1 enzymatic domain. We have summarized an interpretation of our findings in **Figure 4.6**. The model is drawn to scale to reflect total maximum dimension of individual hparp486 domains, DNA, and the hparp486-DNA complexes. We propose that hparp486 interacts with DNA containing overhangs or nicks through an increased area of contact compared to its interaction with blunt end DNA in interactions that lead to the decreased D_{\max} for these complexes. Taking into account the single maximum revealed from the SAXS data, we have modeled this complex as an extended form of hparp486 bound to the area surrounding the DNA damage forming an elongated and asymmetric complex. In conclusion, we believe it is unlikely that PARP-1 is activated by DNA-dependent dimerization, but instead propose a conformational change upon interaction with damaged DNA results in its activation.

Acknowledgement

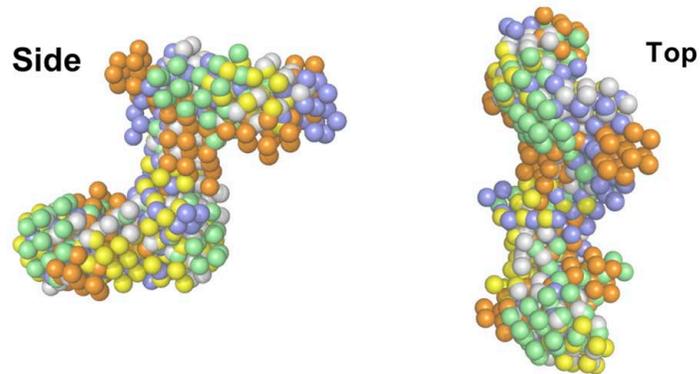
This work was performed by multiple individuals and was so ordered on the peer reviewed Journal of Molecular Biology in which it was published; Wayne Lilyestrom, Mark J. van der Woerd, Nicholas Clark, and Karolin Luger.

Chapter 4

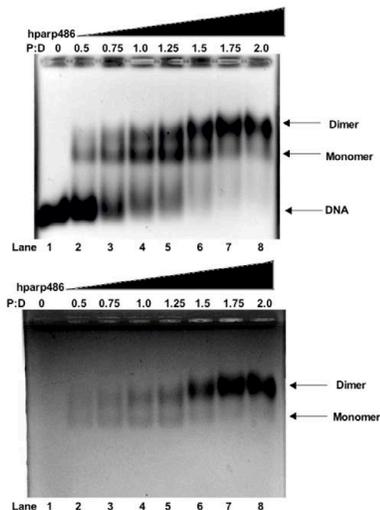
Supporting Information



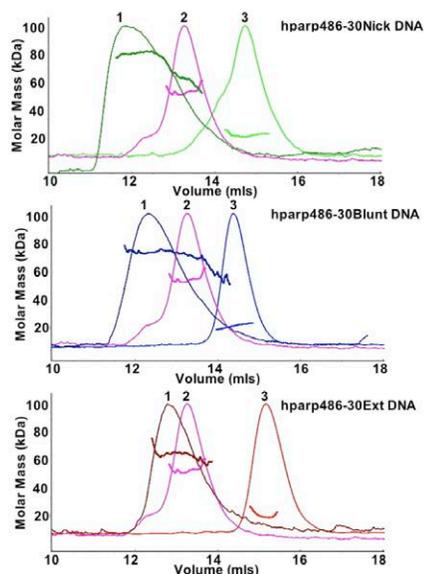
Supplementary Figure 1. parp486 SAXS raw data represented by Guinier, Porod and Kratky analysis. **A)** Superposition of raw parp486 SAXS data at 3, 6 and 9 mg/ml concentrations depicts minimal concentration dependent scattering effects. **B)** Guinier analysis of the low scattering angles depicts minimal aggregation. The residuals are illustrated in the lower portion of the graph. **C)** Kratky analysis of the data illustrates a folded, yet flexible particle. **D)** Porod graph of parp486 SAXS data determined that the sample follows Porod's law and has a volume of $\sim 100,000 \text{\AA}^3$.



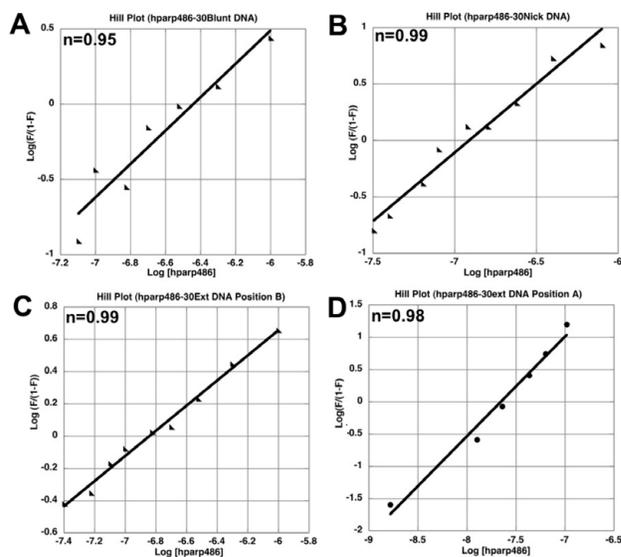
Supplementary Figure 2. Superposition of *ab initio* SAXS models of hparp486. The side (left) and top (right) views of five of the thirty averaged models that were used to create the final hparp486 particle reconstruction are shown. Each model is represented in a unique color. While the models do differ in certain areas, the overall shape of the particle remains similar. These models fit the SAXS scattering profile with a final χ^2 between 1.05-1.15.



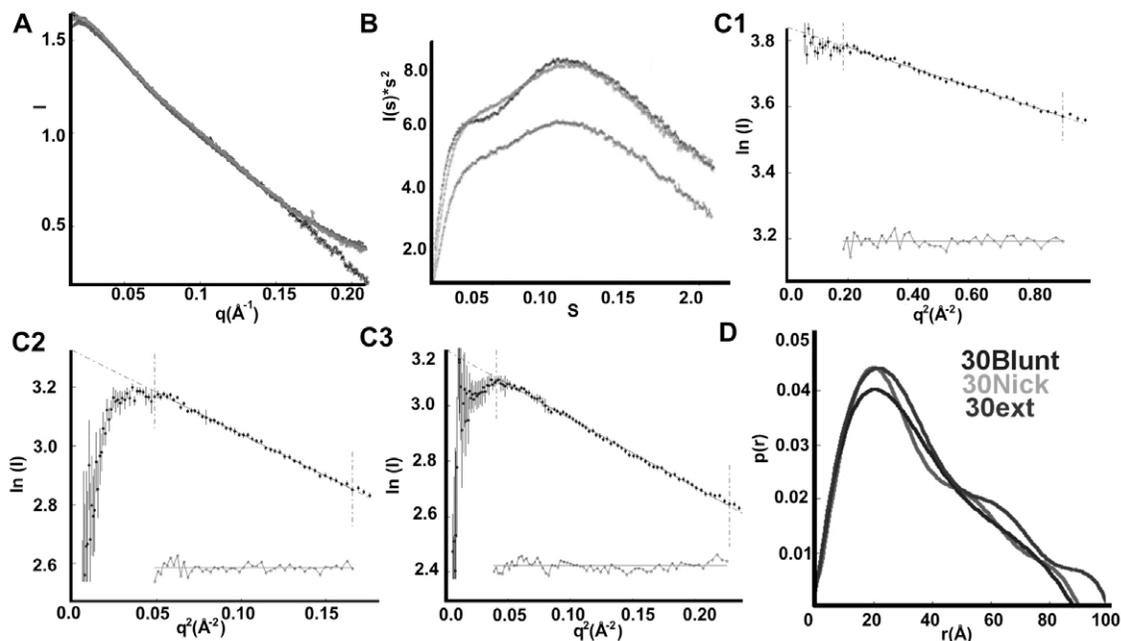
Supplementary Figure 3: EMSA of hparp486 interacting with a 53 base pair blunt ended DNA. Top: EtBr stain; bottom: Coomassie stain. The longer DNA fragment allows hparp486 to interact with DNA in either monomer or dimer form at hparp486:DNA ratios between 0.5-1.75. As the ratio of hparp486 increases, the dimer form becomes more prevalent. The longer DNA differs from the shorter fragments in which only monomeric interactions are detected both by EMSA and in solution using several different assays.



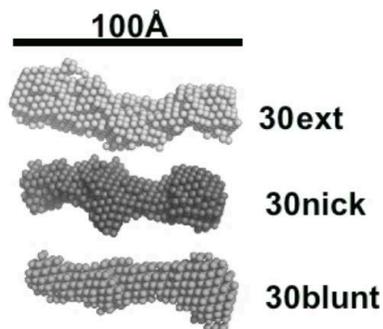
Supplementary Figure 4. Superposition of the SEC-MALS elution profiles of solutions containing either 2:1 molar ratios of hparp486-DNA complexes (peak 1), free hparp486 (peak 2) or free DNA (peak 3). In all cases peak 1 contains a 1:1 mixture of hparp486-DNA as well as free hparp486. This is noted by the decreasing molecular weight in the peak's shallow descent at increased volumes, as well as the significant overlap of the peak 1 with peak 2. The weight average molecular mass is superposed on each peak.



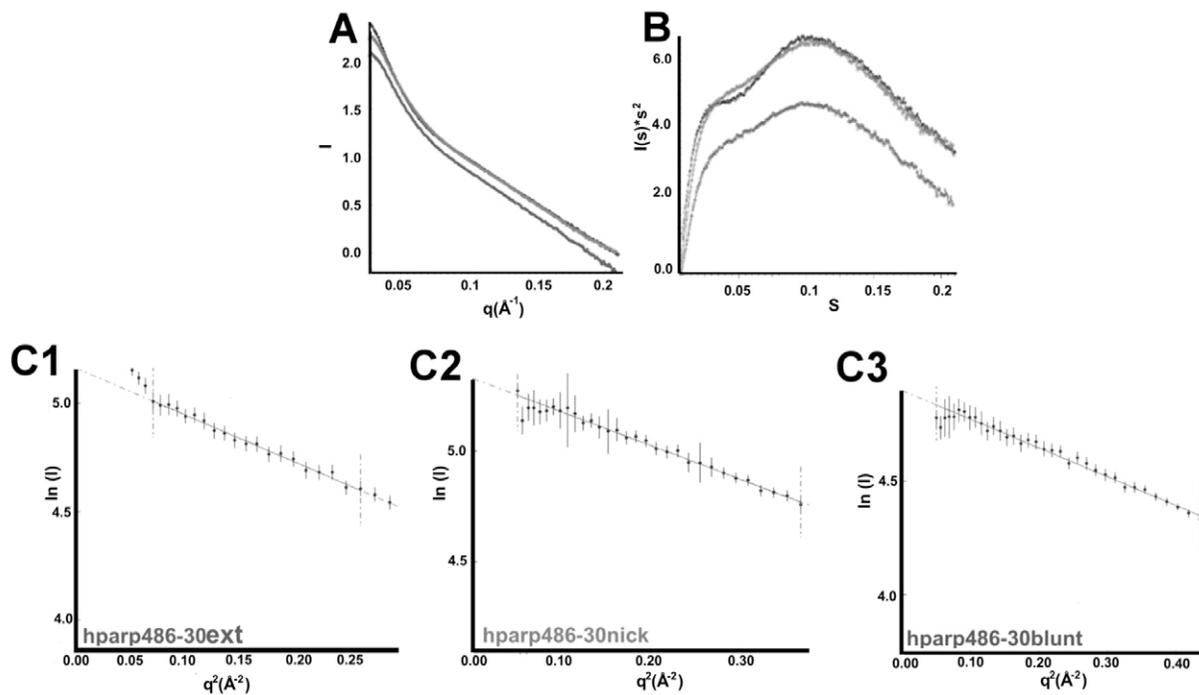
Supplementary Figure 5. Hill plots derived from the interactions of hparp486 with 30blunt, 30nick and 30ext DNA (A, B and C-D, respectively). In all cases, the slope of the line is less than 1 (Table 3). The Hill coefficient is displayed in the upper left hand corner of each plot. Because the Hill coefficient is less than 1.0, these data reflect the lack of positive cooperativity in hparp486 interacting with each DNA damage model.



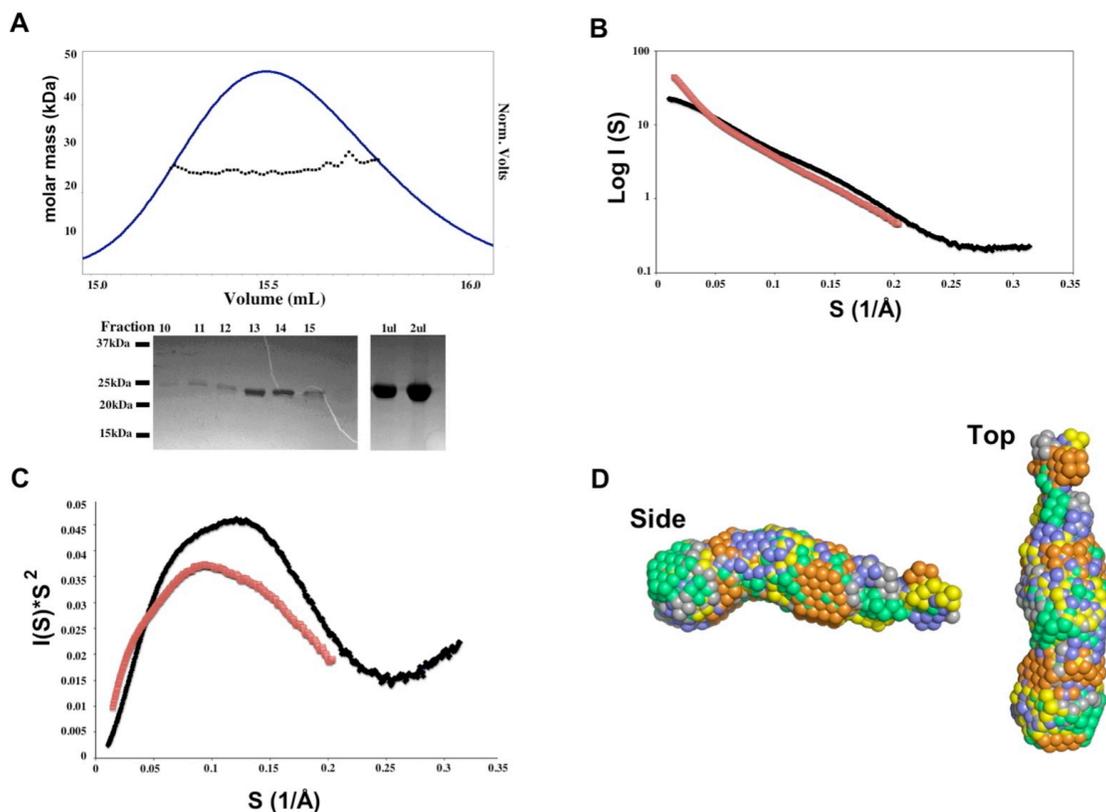
Supplementary Figure 4.6. Analysis of 30Blunt, 30nick and 30ext DNA SAXS data by Kratky, Guinier and $P(r)$. **A)** Superposition of raw SAXS data for 30blunt, 30nick and 30ext DNA. **B)** Superposition of Kratky plots for each DNA illustrate that the 30nick and 30ext DNA do have some long range interactions. **C1-C3)** Guinier analysis similarly depict intermolecular interactions within the 30nick (C2) and 30ext' (C3) DNA samples. Blunt ended DNA (C1) remains homogeneous. The residuals are illustrated in the lower portion of each graph **D)** Superposed distance distribution functions of all DNAs. The data had 2-5 initial points truncated to remove the poor scattering at low angles. The resulting $P(r)$ function portrays that all DNAs are between 88-100Å in the maximum dimension. This result is in agreement with either A or B double helical models.



Supplementary Figure 7. *Ab initio* models of 30ext, 30nick and 30blunt DNA samples. The models were calculated from the average of ten DAMMIN *ab initio* models. The DNA models are all between 88-100Å in length, and cannot be differentiated from either B or A form DNA given the resolution of the technique.



Supplementary Figure 4.8. Analysis of SAXS data for parp486 in complex with blunt, nicked and 6n3' DNA. A) Raw SAXS scattering for each protein-DNA complex. B) Kratky analysis of the complexes reveals an overall similar profile. C1-C3) Guinier analysis of each complex reveals that the data are devoid of aggregation.



Supplementary Figure 4.9. Analysis of Purified hparp209. **A)** Top: SEC-MALS elution profile of purified hparp209 illustrates that the protein is a homogeneous monomer of with a molar mass of 22.5 kDa. **Bottom:** SDS-PAGE analysis of hparp209 from fractions of the SEC-MALS column reveal a singleband at approximately 23 kDa. 1 and 2 μ l of the final concentrated protein fractions that were used for SAXS analysis are shown in the bottom right of panel A. **B)** SAXS scattering profiles of hparp209 (black) and the hparp209-30Blunt complex. **C)** Kratky analysis of hparp209 (black) and hparp209-30Blunt (red) complexes reveals that in both cases the particles have an overall ordered profile, albeit with some disorder at the higher scattering angles in the case of hparp209. **D)** Superposition of five ab initio particle reconstructions of hparp209 that fit the $P(r)$ function (**Fig. 4.5**) with a chi square of 1.1 or better. The reconstructions reveal an over static shape with a small amount of terminal flexibility.

Chapter 5.

The Structural Basis for the Allosteric Activation of PARP-1 by Small Oligonucleotide DNA

Substrates

Chapter Overview

The experiments discussed in this chapter are meant to provide a better understanding of how PARP-1 interacts with DNA-damage. We have extended the work presented in Chapter 3 by incorporating similar studies to the functional full-length enzyme. Advanced small-angle neutron scattering experiments were performed and analyzed with the help of Dr. Susan Krueger and Dr. Joseph Curtis of the National Institute of Standards and Technology Center for Neutron Research (NCNR). The joint PARP-1 project served as a model system to develop methods to analyze flexible proteins, protein-protein and protein-nucleic acid complexes. These experiments would not have been possible without the initial interest of Dr. Wayne Lilyestrom and the willingness of Dr. Karolin Luger to allow me to work independently to maintain the collaboration and to travel to make critical measurements at both the NCNR and the synchrotron SAXS-beamlines at the Advance Light Source (ALS) and Synchrotron Source Radiation Laboratories (SSRL).

Summary:

Scientific evidence in the fields of DNA repair and aging link the protection of the genome to longevity. One protein in particular, Poly(ADP-ribose) Polymerase-1 (PARP-1), has been extensively linked to DNA damage repair. PARP-1 is a DNA damage sensor on its own and is a component of the Base Excision Repair (BER) pathway machinery. DNA damaging events like ionizing radiation or oxidative stress stimulate PARP-1 into action. Upon stimulus, the chromatin-associated PARP-1 catalyzes the addition of polymers of ADP-ribose (pADPr) by modifying its substrate NAD^+ and

covalently linking it to itself and other chromatin associated proteins. It is known that PARP-1 binds damaged DNA at its N-terminus. It is not understood how that DNA-binding event conveys the message that is crucial in activating the C-terminal catalytic domain. We have used a combination of biochemical and biophysical studies to probe the stoichiometry and enzyme efficiency for key types of DNA damage models. Small Angle Neutron, X-ray and Light Scattering experiments were used in a hybrid approach with molecular modeling to reveal structures of the full-length PARP-1 protein, in its entirety, and its complexes with DNA for the first time. The data discussed here support a model in which the binding of damaged DNA conveys a conformational change that is structurally responsible for the activation of the enzymatic function of PARP-1.

Introduction:

The enzyme, Poly(ADP-ribose) Polymerase 1 is the founding member of the PARP superfamily. PARP-1 is highly conserved in eukaryotic organism and it is thought to be the major producer of Poly(ADP-ribose) PAR. PARP-1 is an abundant nuclear protein also known to be involved in transcriptional regulation as well as global regulation of chromatin structure. PARP-1 is crucial for many functions in the cell and its enzymatic activity has been linked to longevity in mammals. Increased enzymatic activity has been correlated positively with lifespan differences amongst various mammals that were tested [90]. PARP-1 catalyzes PAR, a post-translational modification with emerging importance, onto itself and other 'Chromatin Associated Proteins' (ChAPs) [112]. When the cell is in a non-stressed state PARP-1 is thought to bind and help condense chromatin, much like H1. However, many of the structural and functional details for the interaction of PARP-1 with chromatin are unknown. Binding damaged DNA stimulates the catalytic activity of PARP-1 [68]. Following initiation, PARP-1 modifies itself in its automodification domain (residues 374-525) [103]. The modification PARP-1 passes on to itself and other acceptor proteins is a glutamic acid or lysine linked ADP-ribose moiety, which is

synthesized from its substrate NAD^+ molecule. PARP-1 modifies proteins by the addition of multiple ADP-ribose units to make PAR. The discovery of pADPr as a product of NAD^+ catalysis suggests the modification is dependent on a changing nuclear environment [106]. Following the discovery of PARP-1, many other ADP-ribosylating enzymes were discovered all of which share a sequence homology with the C-terminal catalytic domain. Homology to this domain is seen in lower eukaryotes and relegates proteins to the PARP superfamily [23]. However, not all PARP family members reside in the nucleus or share the same activity [224, 225]. Nuclear PARP proteins appear to be highly expressed and PARP-1 has been shown to be the most abundantly expressed of all the ADP-ribosylating family members [123]. Additionally, PARP-1 proteins from different mammalian species exhibit differences in specific activity. This specific activity difference is thought to be a result of variability in primary sequences, which allows for differing levels of automodification. The level of automodification is thought to be one of many factors that correlate with lifespan differences among species [87].

PARP-1 itself is a multi-domain protein that contains several domains in addition to its catalytic domain [124]. The N-terminal region of PARP-1 is the DNA binding domain (DBD), it includes two tandem zinc finger motifs. Adjacent to the DBD resides a third zinc-binding motif, Zn3, this domain has no known DNA binding activity. Connected to the Zn3 domain by a sizable linker is the BRCT (BRCA1 C-terminus homology) domain. The BRCT domain is connected to the C-terminal catalytic domain by an uncharacterized region of secondary structure known as the WGR domain (**Figure 5.1**). Several structures of individual domains have been deposited in the Protein Data Bank (PDB) (**Figure 5.1**) but with its flexible linkers between domains, PARP-1 itself should be expected to be flexible (from fold index, a web-based secondary structure prediction algorithms [226]). The full-length protein is difficult to produce in large quantities in bacteria. One published protocol on bacteria produced PARP-1 discusses a D214A mutant construct that shows a marked decrease in the specific activity of the mutant enzyme [227]. It is thought that the reduction of PARP-1 activity allows for an increase in bacterial production. The difficulty that bacteria have in producing large quantities of PARP-1 may be due to the large size

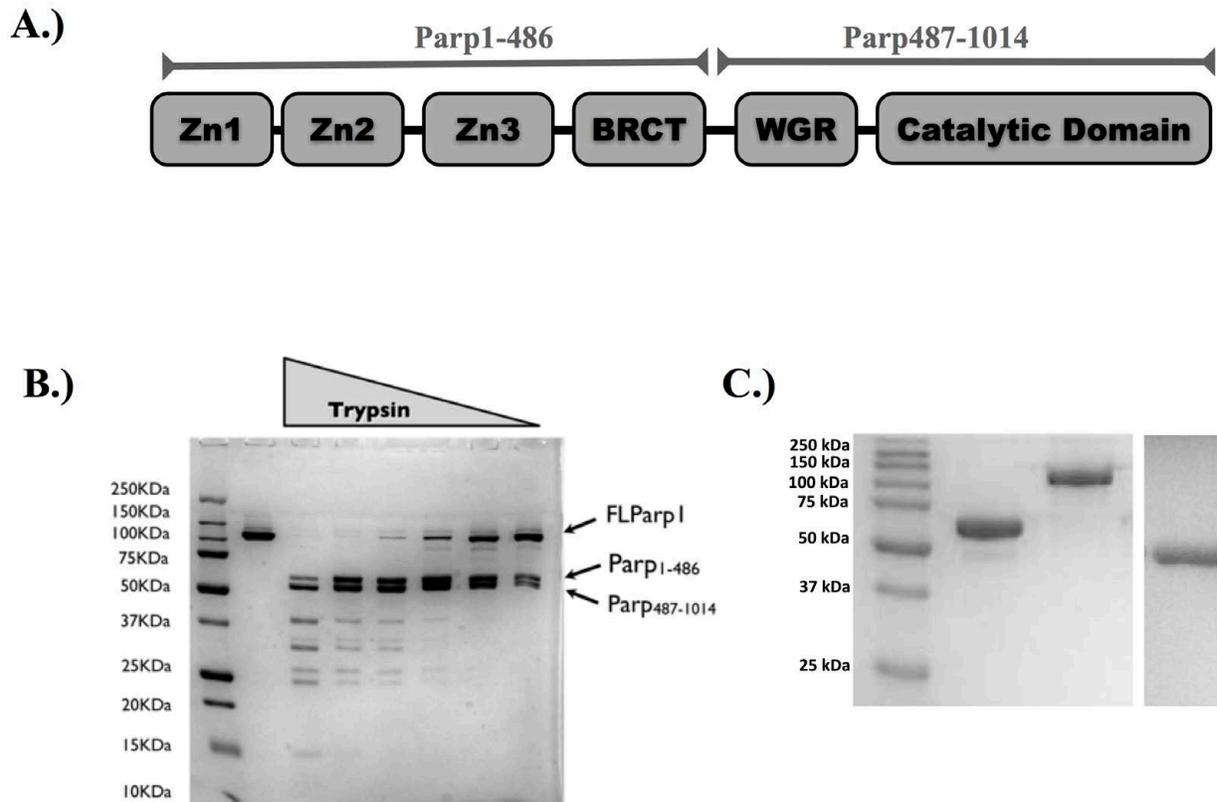


Figure 5.1, The human PARP1 protein contains six known domains. A.) The two N-terminal zinc fingers (Zn1, Zn2) are responsible for binding damaged DNA in a process that activates the C-terminal catalytic domain. Once activated, PARP-1 targets its automodification region, WGR and BRCT domains, and links long chains of ADP-ribose moieties to surface exposed glutamate and aspartate residues. **B.)** PARP-1 was incubated at ratios of trypsin (from left to right roughly 1:100, 1:200, 1:300, 1:400, 1:500 and 1:600) trypsin:PARP for 1hr at room temperature. The reaction was quenched by adding SDS-protein loading buffer spiked with 0.1M PMSF. PARP-1 is primarily cleaved in half. According to N-terminal sequencing, the trypsin cut site is located in the linker region between the BRCT and WGR domains. N-terminal sequencing of the two most concentrated bands, 52 and 56 kDa, shows that the most sensitive site for proteolysis under the conditions used are within the linker region between the BRCT and WGR domains. **C.)** An SDS-PAGE of 3 PARP-1 constructs, from left to right, the His₆-chimera protein Parp1-486, the full-length enzyme containing all 1014 amino acids and the His₆-chimera protein Parp487-1014. These proteins are estimated to be 98% pure. The amino acids that comprise the different halves are denoted by black bars above the protein schematic in **A.**)

(1014 amino acids), its multi-domain structure or possibly toxicity. Bacteria do not have a protective barrier around the cellular DNA, like a nuclear membrane, therefore it is possible that PARP-1 binds bacterial DNA freely. Because PARP-1 becomes enzymatically activated in the presence of DNA it has the potential to deplete cellular NAD⁺ and ATP stores (indirectly) while modifying itself and other proteins with linear and branched chains of ADP-ribose [228]. Bacteria would likely respond to the unregulated hPARP-1 activity by degrading the protein during expression trials (data not shown). To overcome the limitations of bacterial expression of PARP-1 many researchers have resorted to baculoviral expression systems in *Sf9* insect cells [87, 93, 140]. Without understanding how the many domains of PARP-1 interact, its molecular function and the regulation of its activity remain elusive. Understanding of activity is important, for example over-activation of PARP-1 results in necrotic cell death in some instances and survival from cancer in other instances [229]. Thus, the use of PARP-1 competitive inhibitors holds promise during treatment of certain cancer [230].

Many of the individual PARP-1 domain structures deposited in the Protein Data Bank (PDB) do not provide adequate information to understand the overall structure and function; for this understanding the entire (functional) protein is needed. Several of the domains PARP-1 domains in the PDB have little in terms of associated discussions or functional characterization. The biological importance for the individual domains is most likely realized only in the context of the full-length enzyme. As of late, structures of the DNA-binding region and the adjacent domain, Zn1-Zn2 and Zn3 (**Figure 5.1**), respectively, have been published [36, 48]. Important clues to how the Zn1 and Zn2 domains bind DNA have just been revealed. It appears that Zn1 and Zn2 share functional similarities with HMG Zn-fingers and hydrophobic interactions between key amino acids and DNA bases mediate their binding function (**ref**). The structural information and the biochemical characterization of the Zn3 domain suggest its functionality is crucial in the communication of the DNA binding event at the N-terminus to the enzymatic activity at the C-terminus [47, 48]. It is thought that the Zn3 domain can respond to DNA binding at the adjacent Zn1-Zn2-comprised DBD and pass the signal downstream toward the catalytic domain by conformational change [36, 48]. Lilyestrom *et al.* were able to directly show a DNA-induced

conformational change for a fragment of the human PARP-1 protein, suggesting a possible structural basis for the communication between binding and catalytic domains [231]. Recently, a crystal structure of 4 of the 6 PARP-1 domains in complex with DNA was published, and the manner in which they bound DNA was suggestive, to the authors, that a conformational change occurs upon binding DNA [49].

Here we describe the conformational change of the full-length PARP-1 enzyme when bound to small oligonucleotide DNA damage models. We find that PARP-1 is a mono-disperse monomer in solution and it binds DNA in a 1:1 stoichiometry. Small Angle X-ray and Neutron Scattering, contrast variation and global fitting techniques were combined to gain structural insights on several PARP-1 protein constructs and their complexes with DNA. Furthermore, with the crystallography and NMR structural data now available to us [35, 36, 49], we have been able to directly test published atomistic constraints on how PARP-1 interacts with DNA in solution scattering data. Importantly, our studies focused on the intact protein alone and in complex with DNA-damage models similar to those used in the crystallography studies. The structures of the protein and protein-DNA complexes were further evaluated in a modern molecular dynamics program suite, SASSIE [232]. Course-grained Molecular Dynamics allows us to propose the first ever ensemble description of structures present in solution in our studies of the enzymatically functional full-length, human PARP-1.

Materials and Methods:

Cloning, expression and purification of Parp1-486 and Parp487-1014: DNA encoding roughly the first (1-486) and second halves (487-1014) of human PARP-1 was cloned into the PET28a vector system (Novagen). The proteins were expressed (via IPTG induction) in *E. coli* BL21 (DE3) (Novagen) for 6 hours at 25°C. Cell pellets were resuspended in 25 mM Tris pH=7.5, 250 mM NaCl buffer and frozen until use. The protein was bound to a Nickel-NTA fast flow column (Qiagen) and washed with 3 volumes of resuspension buffer (same buffer as above)

prior to being eluted at 300 mM imidazole. Following this, Parp1-486 or Parp487-1014 were further purified by S200 size exclusion and HiTrap SP anion exchange columns (GE Healthcare). To ensure quality control, the purity and homogeneity of the proteins were analyzed by gel filtration, Coomassie-stained SDS-PAGE and static and multi-angle light scattering (Wyatt Dawn Heleos II instrument, ASTRA Software Version 3.2.12).

Expression and purification of full-length PARP-1: Baculovirus with an integrated gene encoding the sequence for human PARP-1 (V762A) was used to make viral plaques in T-flask-bound *Sf9* cells. Here on referred to as Parp1-1014. The virus was diluted in TNM-FH media containing 10% fetal bovine serum (FBS) with a final multiplicity of infection (MOI) of $\sim 10^6$. Plaques were harvested using sterile technique and used in 7-9-day amplifications in a volume of 150mls with a density of 1.0×10^6 cells/mL. The amplified virus was then used as a 100x stock for large-volume protein expressions. Spinner flasks containing $1.0-1.5 \times 10^6$ cells/ml grown in 50% ExCell 420[®] and 50% ExCell Titerhigh[®] (Sigma-Aldrich) were infected for a period of no longer than 72 hours at 27°C. At the time of infection 5 μ M ZnSO₄ and 4% FBS were added to aid in expression levels. Cells were harvested by a 20 minute spin at 1100xg. Cells were sonicated on ice in the presence of Roche Complete EDTA-Free protease inhibitor tablets. The supernatant was cleared by an 18-kRPM spin for 20 minutes. Protamine sulfate was added at 1.0mg/ml and the supernatant cleared by 18-kRPM spin for 20 minutes. Lysates were diluted and loaded onto a 5 ml Cation Exchange column. The purest fractions were pooled and run on a 5 ml HiTrap Heparin column. Pure fractions were pooled, concentrated and run on a S200 size exclusion column.

Catalytic Reactions: Michaelis-Menten enzyme kinetics were set up according to Beneke *et al.* [93]. The reactions contained 0.5 nM PARP-1 and 0.75 nM DNA and brought to a final volume

of 30 μL (before NAD^+ addition) of 10 mM MgCl_2 , 50 mM Tris pH=8. In summary, enzymatic reactions were set up in technical triplicates at 30°C for 30 seconds. Post-reaction, the resulting PAR was cleaved by trichloroacetate and bound to a charged membrane (Zeta-probe[®] BioRad Inc.) through a dot-blot apparatus. The membrane was then blocked and probed with an anti-PAR monoclonal antibody. Next, blots were incubated with Atto647-linked goat anti-mouse secondary antibody and imaged on a Typhoon (GE Healthcare, Life Sciences) fluorescence imager with an excitation wavelength of 488nm and emission wavelength $>610\text{nm}$. Dot-blot were analyzed by ImageQuant 5[®] and the resulting data analyzed using the Michaelis-Menten mode of Prism[®] (GraphPad Prism).

Limited Proteolysis: Purified Parp1-1014 was dialyzed into digestion buffer, (25mM Tris pH=7.0, 100mM NaCl, 5% glycerol). Parp1-1014 was incubated at ratios of trypsin (Sigma-Aldrich) (from left to right roughly 1:100, 1:200, 1:300, 1:400, 1:500 and 1:600, **Figure 5.1B**) trypsin:PARP for 1hr at room temperature. Reactions containing bovine thrombin (Sigma-Aldrich) were performed in an analogous way, except Parp1-1014 was incubated with a varying number of units of protease rather than molar ratios. For DNA titrations, a value of 0.5 units of thrombin was added to each reaction. A 14-17mer DNA (a 3bp 3'-overhang DNA) was titrated into reactions and incubated for 30 minutes prior to adding thrombin enzyme. Adding SDS-protein loading buffer spiked with 0.1M PMSF quenched the reactions. Samples were run on SDS-PAGE and imaged or transferred to PVDF membrane (Millipore) and sent for N-terminal protein sequencing (using the Pick 'n PostTM service offered by ALPHALYSE).

Light Scattering Measurements: Samples of either human Parp1-1014, Parp1-486 or their DNA complexes, each at a concentration of 2mg/ml, were loaded onto an AKTA purifier HPLC system. The SEC-MALS running buffer was as follows in 75 mM Tris pH 8.0, 300 mM NaCl

and 1.0 mM TCEP Samples were characterized over a SHODEX-803 (Shodex) or a Superdex 200 10/30 (GE healthcare) size exclusion column at a flow rate of 0.3-0.5 mLs /minute prior to flowing into an inline Dawn Heleos II (Wyatt Technologies) multi-angle light scattering instrument followed by a OPTILAB rEX refractive index detector (Wyatt Technologies). To obtain hydrodynamic information like the Stoke's radius (R_s) of the molecules, gel filtration standards were used to calibrate the columns (BioRad). A dn/dc value of 0.185, 0.170, and 0.180 were used to determine the concentration of Parp1-486, DNA and the Parp1-486-DNA complex, respectively. Data analysis was carried out with Astra Software version 5.3.4. Reported masses are the mass averaged molecular masses.

Analytical ultracentrifugation: Samples were dialyzed extensively against 25 mM Tris pH 8.0, 200 mM NaCl and 1mM TCEP prior to sedimentation. All experiments were performed in a Beckman XL-I using the absorbance optical system and a 4-hole, AN60-Ti rotor. Sedimentation velocity (SV) was performed in a 1.2 cm, 2-sector EPON centerpiece, while sedimentation equilibrium (EQ) was performed in a 1.2 cm, 6-sector centerpiece. For SV, 400 μ l of a 0.7 Abs_{280nm} of full-length PARP-1, Parp1-486 and Parp487-1014 were sedimented at 50,000 rpm for 4 hours at 20°C, with a radial step size of 0.002cm in the continuous scanning mode. A total of 60 scans were analyzed using the Demeler and van Holde method [233] to yield the diffusion-corrected, integral distribution of S over the boundary $[G(s)]$ within Ultrascan (version 9.4). Fitting of the hydrodynamic properties and molecular mass model was performed within Ultrascan as well.

DNA oligomer preparations: The oligomer sequences for the 21ext primer were 5'-ATCAGATAGCATCTGTGCGGCCGCTTAGGGTTAGGG-3' for the overhang strand and 5'-CCCTAAGCGGCCGACAGATGCTATCTGAT-3' for the shorter strand. Adding or deleting

nucleotides from the 3'-region for the sequences shown above generated additional DNA constructs. Oligonucleotide sequences were ordered from Integrated DNA Technologies IDT[®] DNA (USA). The different oligonucleotide models with either 3'-overhangs or the double-stranded DNA blunt-ended models were created by annealing equimolar amounts of paired primers. To anneal DNAs, the oligomer were mixed and placed in an Eppendorf[®] tube, that was inserted in a 1 liter boiling water bath for 5 minutes, then allowed to cool to room temperature over night. Care was taken to avoid light exposure of labeled oligomers. The homogeneity of annealed DNA was tested by analyzing samples on 20% TBE-PAGE stained with ethidium bromide.

Mobility-shift assays: Protein-DNA complexes were assayed for homogeneity and stoichiometry by analyzing their mobility on 0.7% native agarose gels stained with Coomassie and ethidium bromide to visualize protein and DNA respectively. Low EEO agarose was purchased from Fisher. Native agarose gels were run at 4°C in 25mM Tris pH. 8.0, 20mM glycine, at 60 volts for 60 minutes.

Small Angle X-ray Scattering: SAXS data for the human PARP-1 constructs, DNA constructs and their complexes were measured at the SIBYLS beamline (12.3.1) at the ALS using a Mar CCD area detector (165 mm diameter). A 15µl sample was placed in a 1 mm thick chamber with two windows of 25 µm mica. The detector to sample distance was 1.5 m. A dual double-crystal monochromator was used to select an energy of 10 keV ($\lambda = 1.298 \text{ \AA}$) to optimize the signals corresponding to larger protein complexes. The curves were measured at concentrations of human PARP-1 proteins at 4mg/ml, 2mg/ml, and 1mg/ml; DNA samples at 6mg/ml and 3mg/ml; and human PARP-DNA complex samples equal to 3mg/ml, 2mg/ml and 1mg/ml. Data images were subjected to circular integration, normalization, and subtraction of sample and buffer image files.

All datasets collected were nearly free of concentration-dependent effects, except for the DNA samples with overhangs, which did reveal some evidence of concentration-dependent effects, and had the initial 2-5 data points removed. For DNA, the data sets at lower concentrations were processed for shape determination. The R_g for each particle was approximated using PRIMUS [165] to evaluate the Guinier equation and GNOM [166] to evaluate the $P(r)$ function. The value of the maximum diameter of the particle, D_{max} , was determined empirically by examining the quality of the fit to the experimental data for a range of D_{max} values.

Ab initio particle reconstructions were built by the program DAMMIN [167]. This software represents the protein as an assembly of dummy atoms inside a search volume defined by a sphere of the diameter D_{max} . Starting from a random model, DAMMIN employs simulated annealing to build a model that will fit the experimental data and also fulfills stringent local protein-like conditions, such as chain connectivity and particle compactness. For each final model, thirty independent DAMMIN reconstructions were aligned and averaged with the programs DAMAVER and SUPCOMB to reduce the likelihood that the models represent local minima [234]. The most typical averaged model was filtered to an appropriate volume with the program DAMFILT (*ab initio* models are shown and further discussed in **Chapter 3**).

Small Angle Neutron Scattering (SANS): Similar to the SAXS sample preparation, concentrated Parp1-1014 and Parp1-486 proteins (~4mg/ml) and DNAs (~20mg/ml) were dialyzed in 0, 20, 70, 80 and 100% D_2O complex buffer (300mM NaCl, 50mM Tris pH=8). Protein-DNA complexes were formed after determining the proper ratio on gel shifts. Once large-scale complexes were formed, they were checked for monodispersity in SEC-MALS. To determine any D_2O dependent effects, complexes were monitored by SEC-MALS in the 100% D_2O buffer. Aside from losses attributed to sticking to dialysis membranes, no increase in the visible aggregation, apparent as a small peak in the

void volume of the S200, was present. When losses due to sticking became too great the addition of 5 mM L-arginine alleviated the sticking. Both the buffer and samples were degassed prior to being loaded into 1.0 mm or 2.0 mm lollipop CD cuvetts. The cuvetts were then placed in the beamline at the NCNR's NG-3 30 meter SANS instrument [235]. Samples and their dialysis-reference buffers were measured with the detector at three different distances 1.5 meters, 4.5 meters and 15 meters (the specific distances used were for data collection at the high-, middle- and low-q ranges, respectively) [235].

SANS experiments often require the incorporation of multiple data sets for the same sample in buffers containing different D₂O percentages. The purpose of varying the concentration of D₂O in the buffer is to exploit a key difference in the properties of how X-rays and neutrons interact with matter. Neutrons interact with the nuclei of matter rather than the electrons, which interact with X-rays. Exogenous neutrons, like those in a SANS beamline, demonstrate a significant sensitivity to atomic isotopes and it is this physical property we harness in contrast variation experiments where we use heavy water (the stable isotope deuterium is substituted for hydrogen). Additional benefits for using SANS are that measurements made at different detector distances can be added together so the scattering from different length-scales can be evaluated simultaneously. For instance, measurements made at the 13 meter detector distance will yield information on the largest length-scale, measurements taken at 4.5 meters yield the medium length-scale information while the 1.3 meter SANS and the SAXS data will have the most information on the smallest length-scales (the high-angle data of SAXS is of greater intensity than for that seen in SANS). The ability to incorporate and simultaneously evaluate many parameters makes SANS a very powerful global fitting method. Certain experimental snags or inconsistencies, often due to poor determination of protein concentrations, and intermolecular interactions caused by high concentrations and or D₂O effects [236], can be evaluated in the context of other methods and the most logical interpretation of the data can be determined.

The contrast series we used in our global fits were of samples that were dialyzed into buffers containing 0%, 20%, 70%, 100% D₂O for the full-length enzyme and an additional contrast point, a sample with 80% D₂O in the buffer, for the Parp1-486 construct. The 0% D₂O data was used to evaluate

the protein in the SAXS-like H₂O condition. The other D₂O percentages were used in order to exploit the differences neutron-sensitive molecular properties of the DNA and protein. The atomic composition of DNA and protein differ considerably, and so too does the ability of neutrons to scatter off of those macromolecules. The differences are quantifiable and depending on the molecules' atomic composition, there will be a certain scattering length density (SLD) for which neutrons interact. Since the neutron was discovered many scientists have worked out these SLD values for most of the elements and they can be found on internationally accepted tables. For DNA, the SLD is roughly the same as a solution containing 70% D₂O and 30% H₂O. When a protein-DNA complex is measured in a SANS experiment, using a 70% D₂O buffer, the DNA in that mixture will have the same SLD as the buffer in which it is contained, this is known as the DNA's match point. When DNA is 'matched' the scattering signal that results is from the protein alone. In the 20% D₂O condition, the DNA has a greater signal than the protein, similar to the increased signal from the more electron dense DNA in a SAXS measurement. Since the method depends on global fitting, to solve the contrast variation equation for a system containing two components, one needs to have contrast data at several different SLDs to satisfy the cross-terms of the equation (see Glatter and Kratky [237]). Likewise, the SAXS data for all of protein and protein-DNA complexes was used as additional contrast data since the scattering length density (SLD) for the molecules are different for X-rays.

SANS measurements for an empty cell and an empty beam were made at all detector distances and the scattering profiles were put on an absolute scale by relation to the neutron flux of the beam. Buffer measurements for all detector distances were also collected and the scattering signal was subtracted from that of the sample. Data reduction was performed according to NCNR protocol using IgorPro (Wavemetrics) NCNR SANS data reduction module macros [238].

Molecular modeling: Using X-ray crystal and NMR structural information, we incorporated the known structures of each of the domains of PARP-1 into a starting model that contained atomic coordinates for these domains, a crucial beginning step for molecular modeling (pdb files used 2DMJ,

2CS2, 2RIQ, 2COK, 2CR9 and 2PAW). About ~10% of the protein remained unaccounted for and was assumed to consist of disordered loops, so those amino acids were built with the aid of the program Modeller [239]. The atomistic-DNA model, based off the 21-27mer 3'-overhang sequence, was generated in the "make_na" webserver (<http://structure.usc.edu/make-na/server.html>). For DNA-bound PARP models, the 21-27mer 3'-overhang DNA was first aligned to the Zn2-DNA (pdb 3ODC) structure, then the Zn1 (pdb 3OD8) region was built with the DNA-penetrating loop intercalated upstream from the Zn2 binding region as proposed in the manuscript describing the PARP-1 Zn-fingers binding to DNA [36] (**Supplementary Figure 5.2**).

Starting structures of the proteins and their complexes were energy minimized to determine any improper constraints in the molecular models using the program NAMD [240]. As a means to better understand how the molecules behave in solution, we used a coarse-grained, Monte-Carlo based, molecular dynamics program, SASSIE [232]. Atomistic models were centered and aligned around their centers of mass. The starting models were run through the dihedral generator program where the reference PDB-file was linked to a DCD-file containing the atomic coordinates for the entire ensemble. All atoms were kept rigid except for user designated regions of disorder defined by amino acid corresponding to their position in the protein sequence, (linker region 1 = 1-5aa, linker region 2 = 91-112aa, linker 3 = 202-224aa, linker region 4 = 350-387aa, linker region 5 = 487-528aa and linker region 6 = 640-664aa). For practical purposes, the second linker region for all protein-DNA complexes was held rigid within the dihedral generator program. The dihedral program generated up to 10^4 structures by randomly varying sets of dihedral angles for amino acids within the user-defined flexible linker regions. As the structures were created they were energy minimized using NAMD and the CHARMM-22 force field to ensure steric clashes or overlaps did not occur. Generated atomistic structures were then run through the CRYSON program [[241, 242]. In CRYSON each generated structures was processed and theoretical scattering curves calculated. SAXS and SANS data for the Parp1-486, Parp1-1014 and their DNA complexes were interpolated so that all of their scattering profiles were matched in data points and extrapolated to $I(0)$. Finally, the generated models were run through a filtering program that sorts

structures according to the fit of each scattering profile to the actual data, and values of R_g and χ^2 fits to SAXS or SANS data. The vacuum electrostatics for Parp1-1014 were calculated using the APBS program within VMD [243].

Results:

Successful purification of large quantities of PARP-1 is only achievable by eukaryotic expression methods

Much structural information is available on the individual domains of Poly(ADP-ribose) Polymerase 1 (PARP-1). Very little information on how the individual domains of PARP-1 interact is known. Efforts to produce large quantities of full-length human PARP-1 in bacterial expression systems proved very difficult. We have come to rely upon the baculoviral system for the production of large quantities (~5.0 mg/L) of the full-length PARP-1 enzyme.

After large-scale purifications of PARP-1 were achievable we began to biophysically characterize the enzyme. We found it was a monodisperse monomer in low and high concentration solution studies, analytical ultracentrifugation (AUC) and size exclusion chromatography-coupled multi-angle light scattering (SEC-MALS) (**Figure 5.2**). Interestingly, past publications have described the PARP-1 protein as behaving like a homo-dimer [244]. Because of discrepancies between past publications and our experimental findings we wanted to verify that PARP-1 was fully functional and properly folded. We assayed the activity in both PARP's ability to bind DNA and synthesize poly(ADP-ribose) moieties. We found that Sf9-produced PARP-1 binds and shifts several different types of DNA oligonucleotides in electrophoretic mobility shift assays (EMSA) (**Figure 5.3A**). Consistent with published data PARP-1 only synthesizes Poly(ADP-ribose) (PAR or pADPr) effectively when both NAD^+ and DNA are present (**Figure 5.3B** and **Supplementary Figure 5.5**).

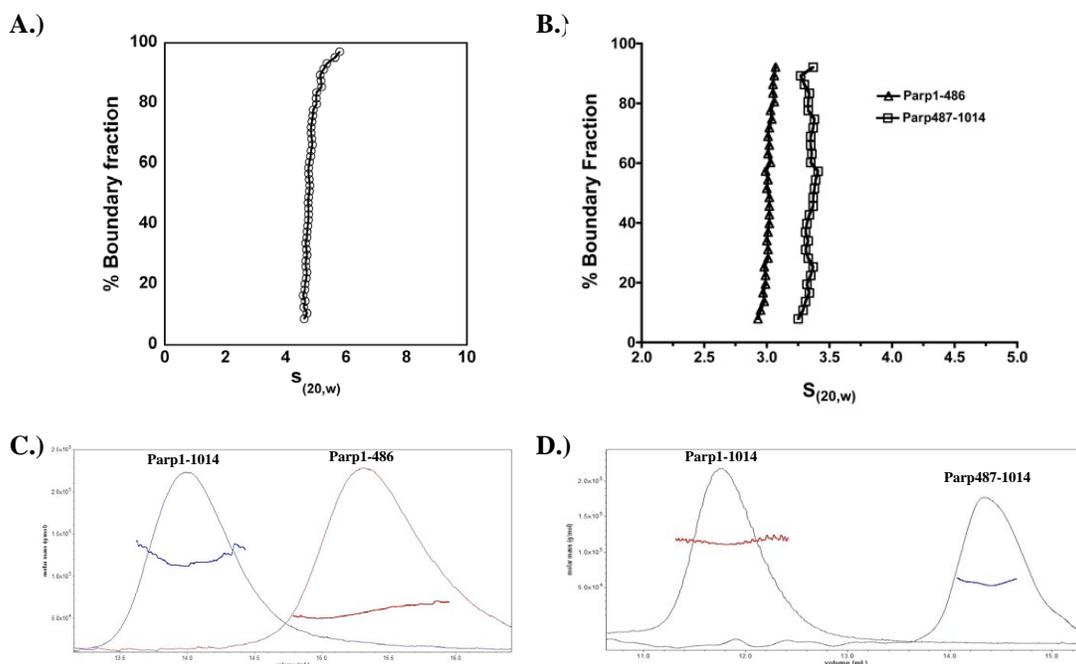


Figure 5.2. Solution studies of PARP-1 indicate it is a monodisperse monomer. **A.)** and **B.)** graphs from analytical ultracentrifugation (AUC) sedimentation velocity experiment indicate that the sedimentation coefficient for the full-length protein is 4.4, a reasonable value for an elongated 113-kDa protein. **B.)** Sedimentation coefficients of two halves of PARP-1, Parp1-486 and Parp487-1014. **C.)** The size-exclusion-coupled multi-angle light scattering (SEC-MALS) of full-length PARP-1 and Parp1-486 indicate molar masses of 111 and 57 kDa, respectively. **D.)** Results from SEC-MALS experiments of full-length PARP-1 and Parp487-1014. The molar masses are 112 and 56.5 kDa, respectively.

Table 5.1. The hydrodynamic properties of the Parp1-1014, Parp1-486 and Parp487-1014. Listed are various hydrodynamic properties derived from the different solution studies of the PARP1 proteins. Stokes radii were obtained by using protein standards of known size and molecular mass within the size-exclusion column attached to the multi-angle light scattering system. Svedberg ($S_{20,w}$) values were obtained for uncomplexed proteins by AUC. For protein-DNA complexes, small amounts of free DNA made sedimentation velocity experiments difficult. Instead, SEC-MALS was used for these protein-DNA complexes. A method to approximate the molar mass for molecules in solution was used and the calculated results for the PARP1 proteins are close to the theoretical values (Erickson, 2009) (*N/A data not acquired).

| Molecule | Elution Volume (mL) | R_s (nm) | Svedburgs $S_{20,w}$ (calculated) | Svedburgs $S_{20,w}$ (measured) | MW (kDa) (theoretical) | MW (kDa) (calculated) | % error |
|--------------------------------|---------------------|------------|-----------------------------------|---------------------------------|------------------------|-----------------------|---------|
| Parp1-1014 | 14.0 | 7.5 | 4.3 | 4.3 | 113.1 | 137.0 | 17.4 |
| Parp1-486 | 15.3 | 5.1 | 3.0 | 3.0 | 56.5 | 64.3 | 12.2 |
| Parp487-1014 | 15.6 | 3.4 | 3.3 | 3.3 | 56.7 | 47.2 | 16.8 |
| 21-27mer DNA | 16.2 | 2.3 | N/A | N/A | 14.8 | N/A | N/A |
| 21-27merDNA:Parp1-1014 Complex | 13.5 | 9.2 | N/A | N/A | 127.8 | N/A | N/A |
| 21-27merDNA:Parp1-486 Complex | 14.6 | 6.8 | N/A | N/A | 71.2 | N/A | N/A |

Limited proteolysis of full-length PARP-1 reveals two stable fragments

Next, we determined if PARP-1 was properly folded by limited proteolysis. We expected that a slow, controlled proteolysis treatment would yield the 6 distinct and fully folded domains that have been described structurally. If the protein was improperly folded, we expected to see extreme sensitivity to protease. Limited proteolysis by trypsin instead yielded two main proteolytic fragments with apparent masses equal to 56 and 52-kDa in size, respectively (**Figure 5.1B**). The result suggests that the protein is properly folded since only a few of the ~150 predicted trypsin cut sites are actually accessible.

Further analysis of the proteolysis experiment, mass spec analysis (MALDI-TOF) and N-terminal sequencing, revealed that the faster moving SDS-PAGE band has a molecular weight of ~56.8-kDa and contains the WGR and catalytic domains as seen by western-blot using a monoclonal antibody raised against that domain (data not shown). The larger species with the apparent 56-kDa size turned out to be similar to the Parp1-486 fragment, containing the Zn1, Zn2, Zn3 and BRCT domains, characterized previously by Lilyestrom *et al.* [231].

The limited proteolysis experiments proved promising for crystallization studies since the experiments yielded two considerably stable fragments. Bacterial expression constructs were successful at producing both the Parp1-486 and the Parp487-1014 fragments at relatively high concentration (6 and 18 mg/L respectively). Although extensive crystallization screens were tried, the two protein fragments did not crystallize. The extensive biophysical characterization of Parp1-486 in Lilyestrom *et al.*, described a well-behaved fragment with functional DNA binding activity [231]. Additionally, the Parp487-1014 fragment is well behaved in solution studies (**Figure 5.2B and 5.2D**) but its catalytic activity was not detectable in enzymatic activity experiments (data not shown). The undetectable level of activity of Parp487-1014 was is not a surprise since the enzymatic activity of the catalytic domain, Parp662-1011, was previously described as being 500-fold less than the full-length PARP-1 [43]. Our biophysical investigations showed that although the Parp1-486 and Parp487-1014 protein fragments are

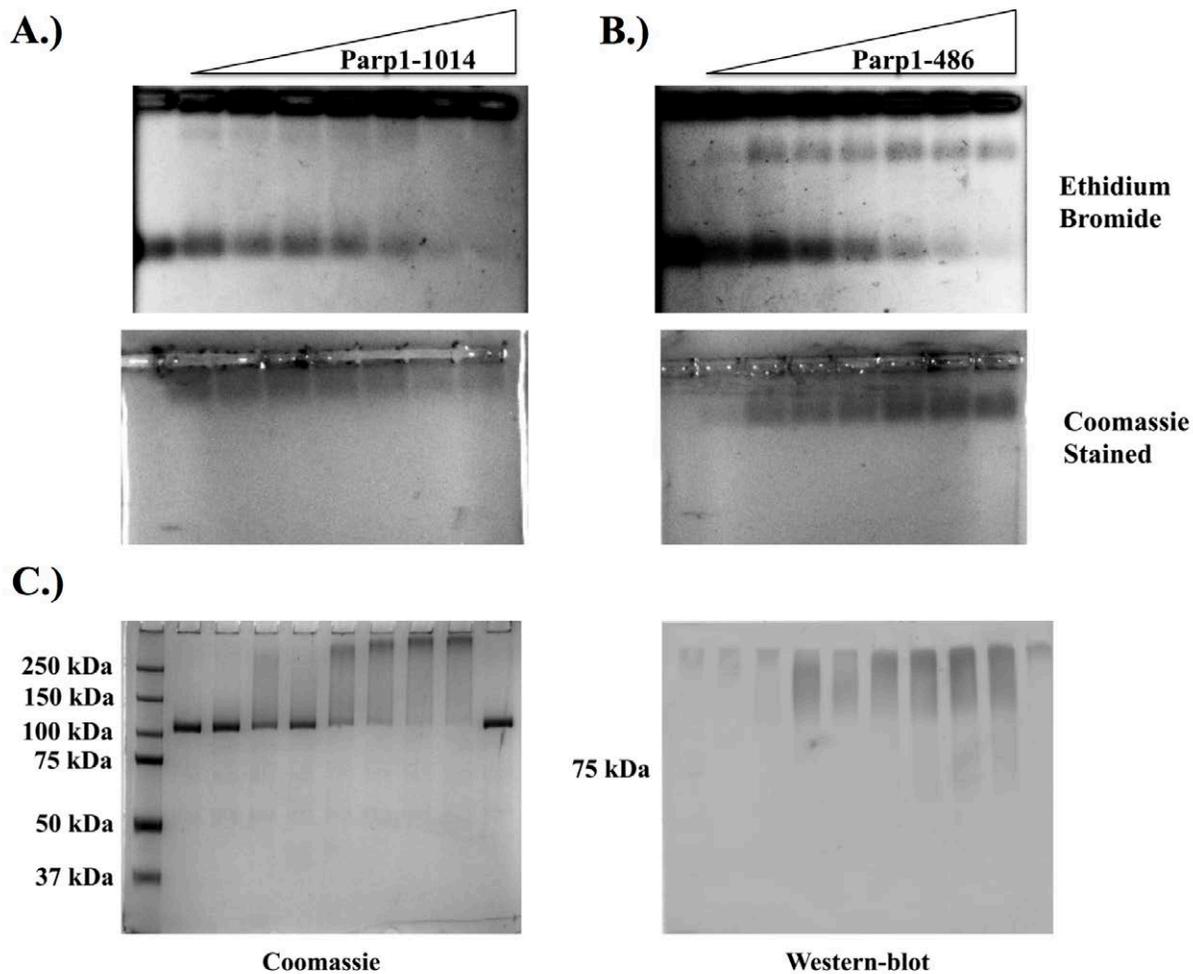


Figure 5.3. PARP-1 is fully functional in the presence of small oligonucleotide DNA-damage model substrates. A.) Electrophoretic mobility shift assays (EMSA) for both PARP-1-1014 and PARP-1-486 indicating binding and shifting of the 21-27mer 3'-overhang DNA. The upper images in A.) and B.) are ethidium bromide stained gels of the PARP-1-1014 (left) and PARP-1-486 (right) titrations to a fixed amount of DNA. Below, in the same order, are the Coomassie-stained versions of the same gels. Judging from migration of the complexes into the gel, it is apparent that the Parp1-1014-DNA complex is less mobile within the native agarose, compared to the Parp1-486-DNA complex. C.) Left is the 8% acrylamide SDS-PAGE gel of a Michaelis-Menten reaction for Parp1-1014. Right is a complementary western-blot analysis run side-by-side under the same running conditions as the Coomassie stained gel. The blot was probed with anti-PAR monoclonal antibody followed by goat anti-mouse secondary antibody, conjugated with an Atto647 fluorophore. Parp1-1014 automodifies itself upon the recognition and binding of DNA, the 21-27mer 3'-overhang DNA damage model. With increasing NAD^+ concentration, an increase in higher molecular weight PAR is seen in the gel and western blot (NAD^+ increases from left to right in a concentration range of $10\mu M$ to $300\mu M$).

of roughly the same molecular mass, 56.5-kDa and 56.8-kDa respectively, the Parp1-486 is more elongated by AUC and size exclusion chromatography (**Table 5.1**).

Structural studies of PARP-1 indicate the full-length protein is more compact than truncation mutants.

The intrinsic flexibility and regions of disorder in the full-length PARP-1 guided our structural studies toward solution-based methods. We began using Small Angle X-ray Scattering (SAXS) experiments to gain low-resolution structural insights. Initial SAXS studies on the Parp1-486 and Parp1-209, published in Lilyestrom *et al.* described the longer larger fragment as being elongated and flexible while the Parp1-209, containing the Zn1 and Zn2 DNA binding domains (DBD), was is more rigid and more condensed in nature [231]. In addition, we wanted to understand how the flexibility of the Parp1-486 and Parp487-1014 fragments compared to the flexibility of the full-length enzyme. Pure (~98% by SDS and SEC-MALS) full-length PARP-1, Parp1-486 and Parp487-1014 proteins were evaluated for monodispersity by SEC-MALS. All three proteins eluted as monodisperse monomers and none showed any real tendency to aggregate (**Figures 5.2 and 5.4**).

SAXS experiments were set up using three concentrations of each protein (usually 4, 2 and 1 mg/ml). Data were then scaled and merged to detect concentration dependent effects like intermolecular attractions (the curves will drastically increase as they approach the Y-axis asymptote) or repulsion (the curves will have a sudden drop in signal as the q-values approach zero) (data not shown). All of the PARP-1 proteins seemed to have only slight intermolecular attraction, by Guinier analysis, in the highest concentration range so, only the lower

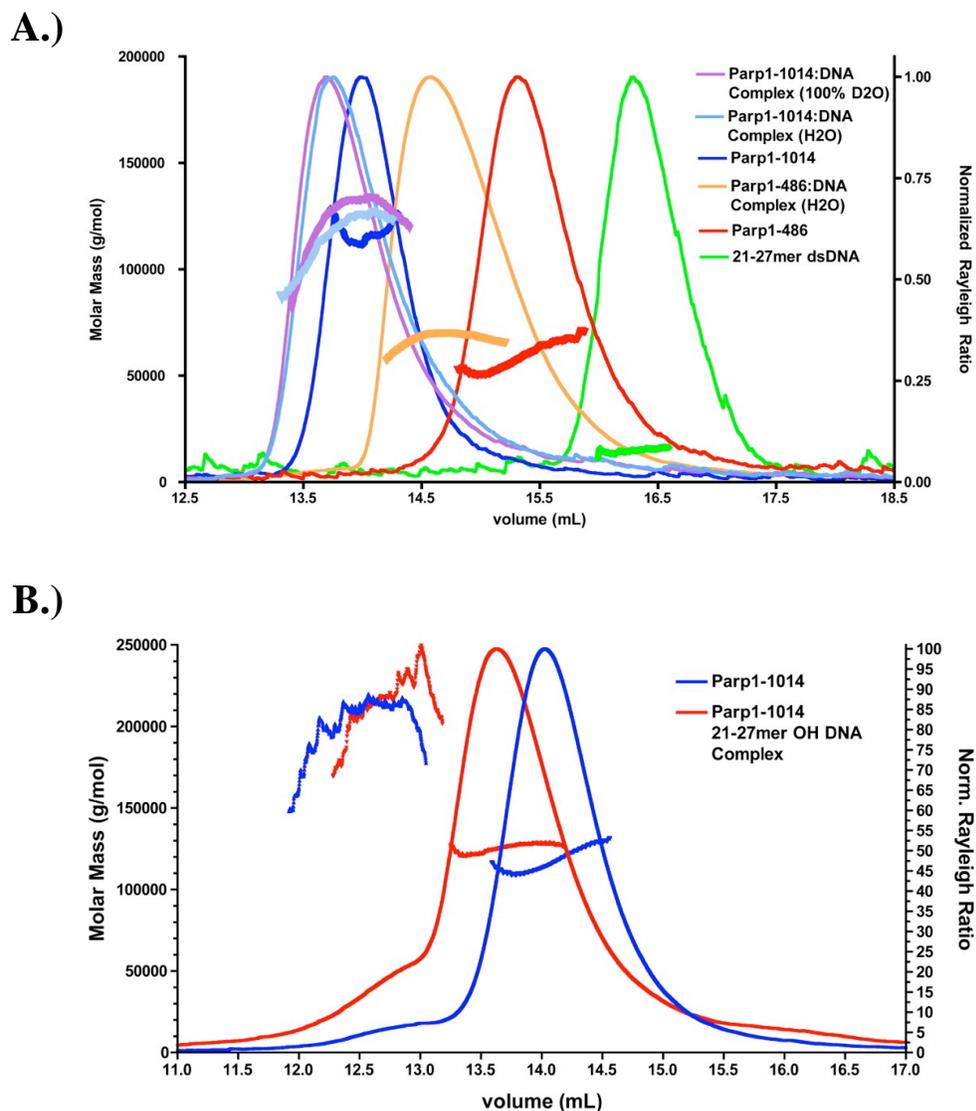


Figure 5.4. Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) of full-length PARP-1 and PARP1-486 and their complexes with 21-27 bp overhang DNA. A.) The PARP-1 and PARP1-486 are highly pure, monodisperse monomers in solution with experimental molecular masses of 111 kDa and 56 kDa, respectively. The experimental masses for the complexes with DNA are 127 kDa and 72 kDa. The mass for DNA is 14.7kDa. Aliquots from these same samples were used for all SAXS experiments discussed here. **B.)** The elution profiles of the PARP1-1014 and the PARP1-1014-DNA complex used for SANS experiments. The molecular weight was measured as 111kDa while the left-hand shoulder on the peak represents 200 kDa, the mass consistent with a homodimer of Parp1-1014 (~2% of total sample volume). While the Parp1-1014-DNA complex had a mass of ~129 kDa the dimer shoulder remained present and increased to approximately 6% of the sample volume loaded. To prevent sticking of PARP to the membrane, L-arginine was incorporated in the dialysis and size-exclusion buffers. The L-arginine increased the efficiency of dialysis, however, low levels of dimerization were detected in SANS samples of Parp1-1014.

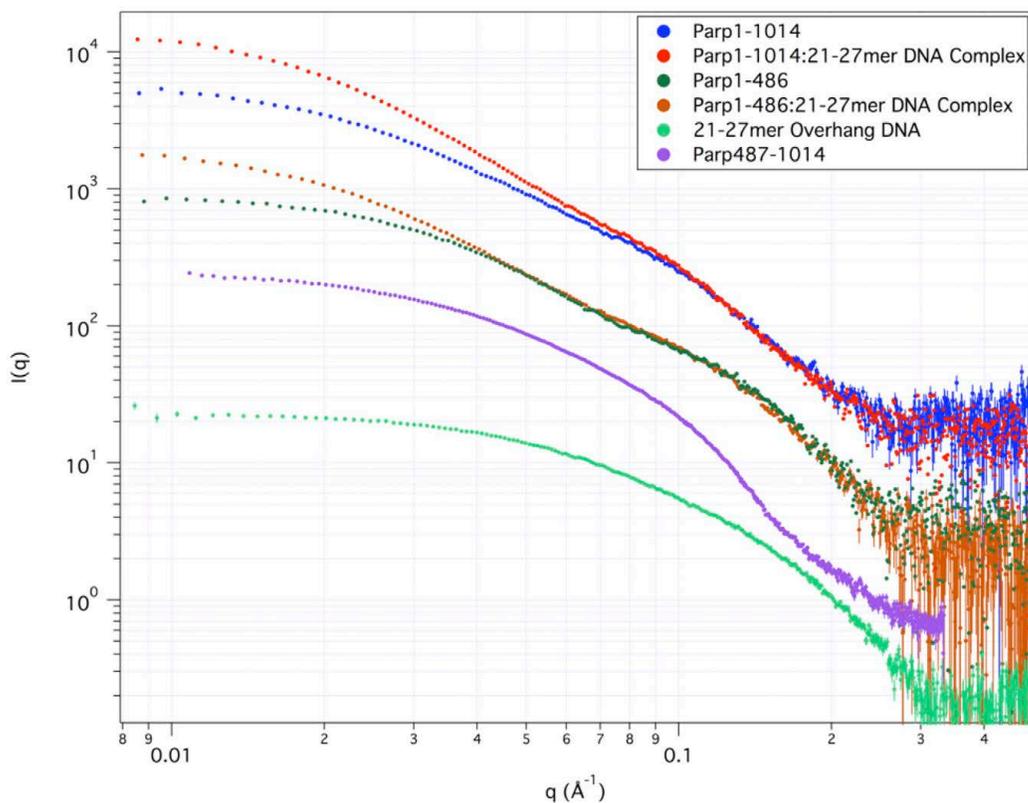


Figure 5.5. Small Angle X-Ray Scattering (SAXS) profiles of PARP-1 and the 21-27mer 3'-Overhang DNA. A.) The scattering profiles from top to bottom: PARP1-1014-DNA Complex, PARP1-1014, PARP1-486-DNA Complex, PARP1-486, PARP487-1014 and the 21-27mer 3'-overhang DNA. The data look clean and remain robust until a q -value of ~ 0.3 the data become too noisy beyond that q -range. None of the samples showed any sign of radiation damage or concentration dependent effects, like intermolecular attractions or repulsions. The data were analyzed using the ATSAS software suite including GNOM and PRIMUS (ATSAS EMBL).

Table 5.2. Results from Size-Exclusion-Coupled Multi-Angle Light (SEC-MALS) and Small Angle X-ray Scattering (SAXS) experiments. The calculated molar masses of the PARP1 proteins and their complexes with DNA match, with great certainty, the measured values obtained in SEC-MALS experiments. D_{\max} values for each species were approximated from $P(r)$ distribution analysis using PRIMUS. For Parp1-486 the values closely match those in Lilyestrom *et. Al* (see Chapter 4). *Only SAXS measurements taken.

| Molecule | Molecular Mass Calculated (kDa) | Molecular Mass from SEC-MALS (kDa) | R_g from SAS (Å) | D_{\max} from Pair Distribution (Å) |
|--|---------------------------------|------------------------------------|--------------------|---------------------------------------|
| Parp1-486 | 56.48 | 54.4 ± 0.5 | 44.0 | 154.0 |
| Parp487-1014 | 56.78 | 56 ± 0.1 | 39.0* | 134.0 |
| Parp1-1014 | 113.08 | 111.5 ± 0.9 | 58.0 | 187.0 |
| Parp1-486:21mer Blunt DNA Complex | 69.28 | 68 ± 1.0 | 67.9* | 210.0 |
| Parp1-486:21-27mer Overhang DNA Complex | 71.24 | 72 ± 0.8 | 67.0 | 205.0 |
| Parp1-1014:21mer Blunt DNA Complex | 125.88 | 124 ± 2.0 | 76.0* | 251.0 |
| Parp1-1014:21-27mer Overhang DNA Complex | 127.84 | 125 ± 1.0 | 79.0 | 248.0 |
| 21mer Blunt DNA | 12.85 | 12 ± 5.0 | 21.0* | 68.0 |
| 21-27mer 3'-Overhang DNA | 14.76 | 14.1 ± 6.0 | 23.0 | 73.0 |

concentrations were used for further analysis (data not shown). Further analysis on the scattering curves was performed to determine the degree of protein folding the using a Kratky plot and a Pair Distribution Fourier transform was performed and the results plotted as a P(r)-plot in order to independently determine the R_g . This method serves as an alternative method to derive the R_g -value and it can be achieved without having to rely on the Guinier region, a region very sensitive to aggregation (**Figure 5.6**). Porod plot analysis was then performed and the volume under the curve calculated for each protein ($q^4 \cdot I(q)$ vs q). From the volume calculations the average molecular mass of the molecules could be calculated and all were very close to their theoretical molecular masses (**Table 5.5**). The scattering curves were then used to calculate *ab initio* molecular envelopes (*ab initio* envelopes are described and shown in **Chapter 3**) and key values like the average radius of gyration (R_g) and the maximum dimension (D_{max}) of the particle (**Tables 5.3** and **Table 5.4** and **Figure 5.6A** and **5.6B**). Both separate halves of the PARP-1 had very similar R_g values while full-length PARP-1 was expectedly larger. A small contribution to the R_g -values for both the Parp1-486 and Parp487-1014 no doubt comes from the 16 N-terminal amino acid residues that remain on the proteins as a cloning artifact from the introduction of a His₆-affinity tag (previous studies suggest the two proteins are very sensitive to proteolysis and the removal of the affinity tag was deemed too risky). As an additional evaluation of the elongated nature of the proteins the ratio of the R_g/R_s -values were calculated according to the work of Tande and Wagner (**Table 5.1**) [245]. Taken together, the data suggest that the full-length and the two protein halves considerably deviate from the shape of a sphere. The possibility exists that an elongation of the Parp1-486 and Parp487-1014 protein fragments may be due to the loss of key domain interactions that are abrogated when the protein is cleaved.

Since we have access to both Parp1-486 and Parp487-1014 constructs, we tried to determine if those domains, now on separate protein fragments, can interact in *trans* to form an active full-length protein. We set up complexes at concentrations in the μ M range but were unable to form a complex, as seen by size exclusion chromatography, SEC-MALS and AUC (**Supplementary figure 5.1**). We

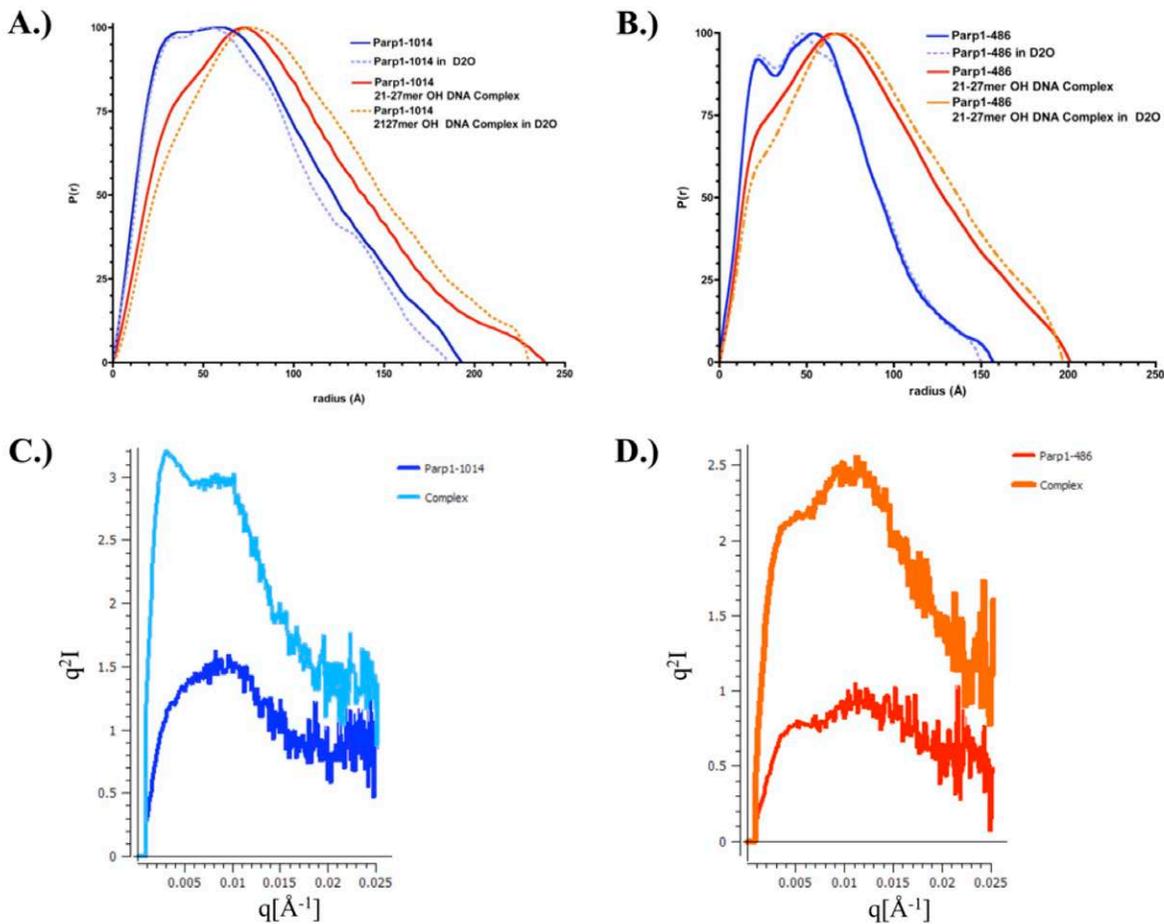


Figure 5.6 Small Angle X-ray Scattering (SAXS) data analysis. A.) and B.) the pair distribution ($P(r)$) plots of Parp1-1014 and Parp1-486, respectively. The blue curve represents the $P(r)$ distribution of the protein alone and the red represents the $P(r)$ distribution for the protein-DNA complex. The overall shape and the distribution of interacting species are very different for both proteins when they are in complex with DNA. Additionally, the maximum dimension (D_{\max}) for both molecules increases by $\sim 60\text{\AA}$ and $\sim 50\text{\AA}$ for PARP1-1014 and PARP1-486, respectively in complex with DNA. To determine the best D_{\max} value, several approximations were made before finding the one that had the best fit to the data. A good fit was in part verified by producing R_g -values close to those calculated from the Guinier analysis. The hashed lines in both A.) and B.) represent the $P(r)$ plots for the molecules in $\sim 100\%$ D₂O buffer. C.) and D.) the Kratky analysis for Parp1-1014 and its complex with DNA and Parp1-486 and its complex with DNA, respectively. From the Kratky plots it is evident that there is a plateau in the high- q region ($q=0.25$). The bell-shaped region signifies that the molecule is well behaved and mostly folded and the plateau region can give information on the degree of flexibility. Because roughly 10% of the protein is comprised of flexible linkers, the protein behaves like it has some regions of disorder (a globular protein would have a plateau that approaches the x-axis). When either protein is in complex with DNA the shape of the curves change considerably and both have a similar increase in the height of the plateau region. One interpretation for the change in plateau height might be a change in flexibility when DNA is present.

determined the ability of Parp1-486, alone and in the presence of DNA, to activate the enzymatic function of Parp487-1014. In those experiments, Parp1-486 was not able to activate Parp487-1014 in *trans* (data not shown). Interestingly, the previously mentioned cut-site for trypsin is located in the linker region between those domains. It seems that without the covalent linkage between the BRCT and WGR domains, the ability of those domains to interact is lost. It is possible that one such linker region, the one located between the BRCT and WGR domains, is integral in relaying the DNA binding event at the N-terminus to the catalytic domain at the C-terminus. The very same linker region was recently shown to contain the three lysine amino acids that serve as critical automodification targets [42]. Automodification, in the form of large chains of negatively charged ADP-ribose, might serve to disrupt potential domain-domain interactions that might be contributing to the more condensed nature of the Parp1-1014 as compared to the more elongated Parp1-486 or Parp487-1014.

Small oligonucleotide DNA damage models activate PARP-1

In Lilyestrom *et al.* we determined reported that 30-base pair (bp) blunt-ended, nicked and overhang DNAs were able to stimulate the full-length enzymatic activity ~40-fold more than that of the enzyme in the absence of DNA [231]. This was an interesting important result because most past previous experiments relied on sheared salmon sperm DNA, most likely a mixture of all kinds of DNA-damage substrates, to activate the enzyme. To better understand the enzymatic parameters of PARP-1 in the presence of different DNA damage models, we performed Michaelis-Menten kinetics assays and compared the relative enzymatic activity levels to the relative binding affinities of PARP-1 to multiple DNA damage models (see **Chapter 2**).

To answer the question whether the activation of PARP-1 leads to an intramolecular modification or an intermolecular modification (self modification vs. modification of a neighboring molecule) we performed a DNA titration assay using the 30mer nicked DNA. Our hypothesis was that PARP-1 acts as an intramolecular modifying enzyme. If indeed the reaction is intramolecular, the reaction rate will

depend only on the DNA concentration, and *vice versa*. If DNA is limiting within a reaction then the number of DNA-activated enzyme molecules should also be limiting. Under such limiting conditions, a molecule that modifies its neighbor should display nearly the same level of activity regardless of DNA concentration (as long as the DNA concentration is greater than its dissociation constant) because the modification substrate is in great excess, lots of free enzyme. If however, the reaction occurs by the enzyme modifying certain residues on its own side-chains, or that a DNA-induced conformational change, then limiting the DNA concentration within the reaction might result in far less activity. The result of the titration experiment indicated that up until a certain concentration of DNA is reached, a ratio of 0.2:1 DNA:PARP-1, the enzymatic activity level is far reduced (**Figure 5.7**). When DNA is limiting (only a given number of molecules can become active-enzymes) and both the substrate (PARP-1) and its enzymatic cofactor (NAD^+) are in excess, the reaction does not reach the maximum capacity for those conditions. When DNA rises to a certain concentration, the maximum capacity of the enzyme reaches a plateau that remains stable even when excess DNA is present (**Figure 5.7C and 5.D**). These results are interesting and might serve as a model for the design of future enzymatic studies. Perhaps using advanced enzymatic techniques and equipment like a stopped-flow mixing apparatus, some of the key properties of the complicated PARP-1 enzymatic parameters can be elucidated.

For the remaining studies discussed here only two of the most stable DNA-damage models were assayed in order to determine differences in enzymatic efficiency: a double-stranded blunt-ended DNA and a double-stranded DNA substrate with a 3'-Overhang (data for the 21mer blunt DNA is not shown). Catalytic reactions where PARP-1 itself served as the modification target were set up with a series of reactions containing increasing $[\text{NAD}^+]$ and molar excess of DNA. The level of automodification present for each reaction was visualized using an 8% acrylamide PAGE-gel followed by a western-blot analysis while probing with an anti-PAR monoclonal antibody (**Figure 5.3**). The enzymatic activity levels were quantitated as described in Chapter 2. Although we saw little difference in the level of activation between DNA damage models, a direct comparison of the enzymatic efficiency between DNA damage substrates has translational importance. Understanding the enzymatic functions PARP-1 in the presence of different

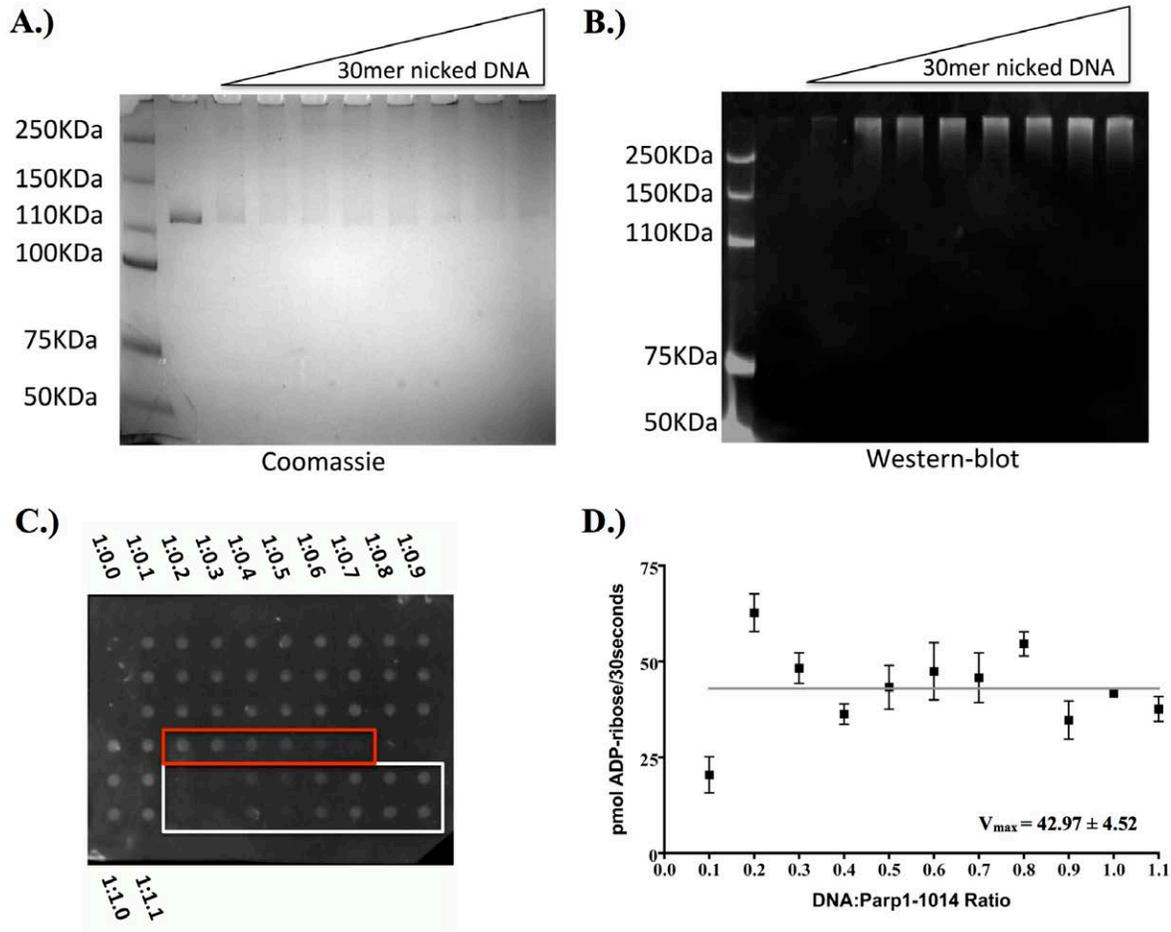


Figure 5.7. The PARP1 enzymatic activity occurs through an intramolecular mode of automodification. **A.)** An 8%-acrylamide SDS-PAGE gel showing an enzymatic reaction where all reaction components except DNA were kept constant. From left to right, an increase in automodified-PARP1 is visible in the high molecular weight smearing, which occurs simultaneously with the disappearance of the 113kDa PARP1-1014 band. **B.)** A western-blot analysis of an identically loaded gel. The blot was first probed with anti-PAR monoclonal antibody followed by goat anti-mouse secondary conjugated with an Atto647 fluorophore. **C.)** Dot-blot with Zeta-probe bound PAR. The PAR was cleaved by addition of 20% TCA and 5% of the total reaction was bound to the membrane. The blot was probed in a procedure similar to the western-blot. The red box denotes the standard curve where a dilution series of known concentration of commercially available PAR was bound to the membrane. The standard curve was used to quantitate the PAR concentrations in the experimental conditions. The white box represents a control in the form of a Michaelis-Menten reaction of a 30mer nicked DNA (the [NAD⁺] ranged from 0-300 μ M). **D.)** A plot of PAR formed in 30 seconds as a function of different DNA ratios used in the reactions. The average V_{max} , of 42.97 ± 4.52 pmol PAR/30sec was calculated from the best-fit line (grey). A plateau in the activity is reached when the PARP:DNA ration is greater than or equal to 0.2:1.0. Below a certain level of complex formation, the enzyme appears to be minimally activated. Perhaps it is not until a threshold of DNA is reached that enough enzyme can be activated to produce a detectable level of automodification. A DNA binding induced conformational change may be required for an efficient activation of the protein.

DNA-damage models is important because specific types of chemotherapeutics are known to cause specific types of DNA damage. We continued our structural studies of PARP-1 with the two different DNA-damage models in order to rule out the possibility that one particular damage type could induce more of a conformational change on the enzyme than another damage type (i.e. a double-strand break site vs. a site with a 3'-overhang).

Structural studies of the complexes between PARP-1 and blunt and 3'-overhang DNA damage models.

Extensive biophysical characterization resulted in our model that suggests Parp1-486 has the ability to undergo a conformational change in the presence of DNA in Lilyestrom *et al.* [231]. It was from this work that we hypothesized that the DNA binding-induced conformational change might be even more significant in the context of the full-length protein, the relevant and functional enzyme. Such a conformational change may be required for the efficient activation of the C-terminal catalytic domain upon DNA-binding at the N-terminus a distance spanning a distance of up to several hundred amino acid residues and, prior to this study, an unknown angstrom length.

Understanding the conformational change in the PARP-1 protein upon DNA binding became paramount. As a means to reduce any possible contribution the DNA might have on particle length, shorter DNA-damage models were designed (i.e. 21mer blunt vs the 30mer blunt used in past studies). As mentioned before, our studies required the most stable DNA damage models possible. To achieve stability we stuck to the simplest DNA damage models, an overhang and a blunt DNA (the nicked DNA consisted of annealing 3 single stranded DNAs, two of which had very low melting points and its use would risk introducing more heterogeneity into any system studied). The 3'-overhang and blunt models consisted of a 21-27mer 6 bp 3'-extension and a 21mer double stranded DNA (dsDNA), respectively.

The major proteins of interest in the current study are Parp1-1014 and Parp1-486, which were used to form complexes with both DNA-damage models. Both proteins are able to form complexes and

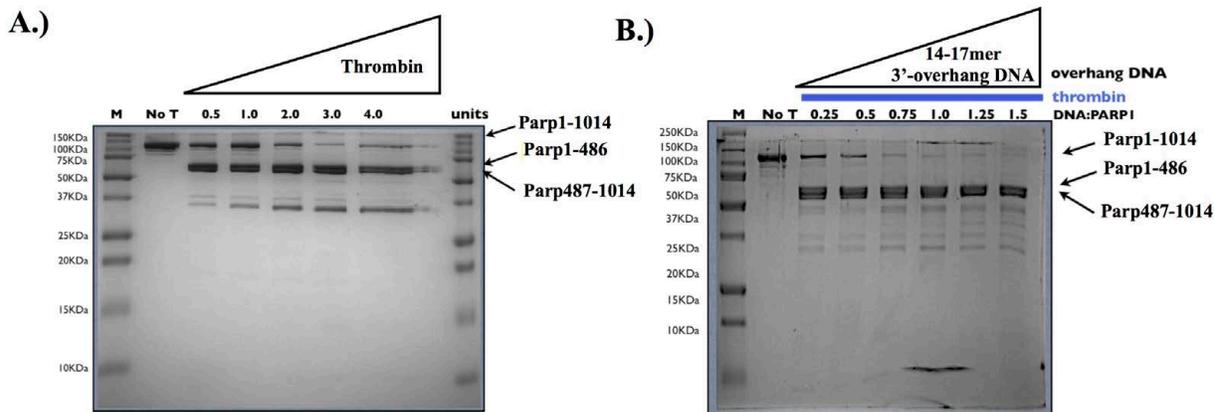


Figure 5.8. Limited proteolysis of PARP-1 in the absence and presence of DNA.

A.) Full-length PARP-1 (Parp1-1014) was exposed to limiting concentrations of thrombin protease. The resulting pattern of digestion was distinctly different than the pattern seen in the presence of trypsin protease (see Figure 1). **B.)** The 0.5 unit concentration of thrombin in A.) was used in a limited proteolysis experiment where DNA was titrated at different concentrations in different reactions. Rather than protecting the protein from digestion the DNA increased the sensitivity of Parp1-1014 to proteolysis. Additionally, having DNA present caused a different digestion pattern, indicating a conformational change in the PARP-1 structure as seen by a different digestion pattern in the presence of DNA. An identical gel to that seen in **B.)** was run simultaneously and transferred to PVDF membrane. The two major bands in the gels were excised and sent for N-terminal sequencing. N-terminal sequencing confirmed the existence of a thrombin cleavage site in the linker region between the BRCT and WGR domains.

shift DNA in EMSA experiments. However, DNA complexes with the full-length protein have a noticeably different mobility within the gel and might represent a greater degree of elongation than that of the Parp1-486-DNA complex (**Figure 5.3**). Large-scale complexes were set up according to gel shift ratios and analyzed by SEC-MALS. Both proteins formed one-to-one complexes equally well with both DNA models. Monodisperse complexes were then used in SAXS experiments at three different concentrations (3mg/ml, 2mg/ml and 1mg/ml). Again, SAXS scattering profiles for the three different concentrations were scaled and overlaid to determine any concentration dependent effects. Only at the highest concentrations did the complexes with full-length PARP-1, and to a lesser degree Parp1-486, show signs of intermolecular attractions (see **Tables 5.3** and **5.4**).

Once the SAXS data were analyzed it became obvious that both full-length and Parp1-486 underwent large, DNA-induced conformational changes (**Tables 5.2** and **5.3**). The values obtained for Parp1-486, these findings are consistent with the published work of Lilyestrom *et al.* (see **Chapter 3**), even though the DNA-damage models used in this study are considerably smaller [231]. Analysis of the complexes by the methods of Kratky, Porod and Guinier were performed. From these methods, we saw that the complexes were well behaved and did not show a tendency to aggregate and the Porod volumes correspond with expected molecular masses (Kratky plots and Porod volumes are shown in **Figure 5.6B** and **Table 5.5**, respectively). We also saw that the complexes underwent a conformational change as seen by P(r) analysis and drastic changes in R_g and D_{max} were observed compared to the protein alone (**Figure 5.6A** and **Table 5.3**).

Small Angle Neutron Scattering (SANS) studies indicate proteins undergo elongation in the presence of DNA.

Macromolecules exposed to X-ray radiation have scattering intensities that are proportional to their electron density. In the case of protein-DNA complexes the electron density for the two different macromolecules is quite different and the contribution of each in a complex can be determined by sucrose

gradient work. A superior method is the application of small angle neutron scattering (SANS). Neutrons interact with atomic nuclei rather than with electrons. Neutrons exhibit a high degree of sensitivity toward isotopes, especially when hydrogen atoms are substituted for deuterium because of the difference in their scattering length densities (SLDs).

Contrast variation and SANS experiments in general require large quantities of material and are very time-intensive. Due to a relatively low intensity beam and the weak interactions with matter, experiments can span several days. The goal of an experiment is to form concentrated (4-6 mg/mL with several mLs of sample needed), monodisperse complexes in various concentrations of D₂O. Often times, D₂O can cause intermolecular interactions that are often exacerbated with increasing concentration of the protein or protein-DNA complex. These interactions are well-known [236] but not well understood. Concentrated Parp1-1014 and Parp1-486 protein fragments (~4mg/ml) and DNAs (~20mg/ml) were dialyzed in ~100% D₂O complex buffer (300mM NaCl, 50mM Tris pH=8). To determine any D₂O dependent effects on the molecular state, complexes were monitored by SEC-MALS in the same D₂O buffer. Occasionally only slight aggregation was apparent in the void volume peak on the chromatograms (data not shown). Protein-DNA complexes in ~100% D₂O were formed for Parp1-486 at 3.0 mg/ml, 2.0mg/ml and 1.0mg/ml and full-length PARP-1 at 2.0mg/ml. Complexes for the 21mer blunt DNA were assembled for Parp1-486 at 3.0, 2.0 and 1.0mg/ml and Parp1-1014 at 2.0mg/ml. SAXS data were collected, scaled and superimposed to determine D₂O-dependent effects. Only slight D₂O dependent effects were seen in the lowest scattering angles with the highest concentrated samples (**Table 5.3** and **Table 5.4** for the Parp1-486 and Parp1-1014, respectively).

SANS analysis of Parp1-486 and the DNA-Parp1-486 complexes

SANS experiments were prepared as before with the Parp1-486 protein and its complex with the 21-27mer overhang DNA. SANS studies of complexes made with the 21mer blunt DNA would not fit

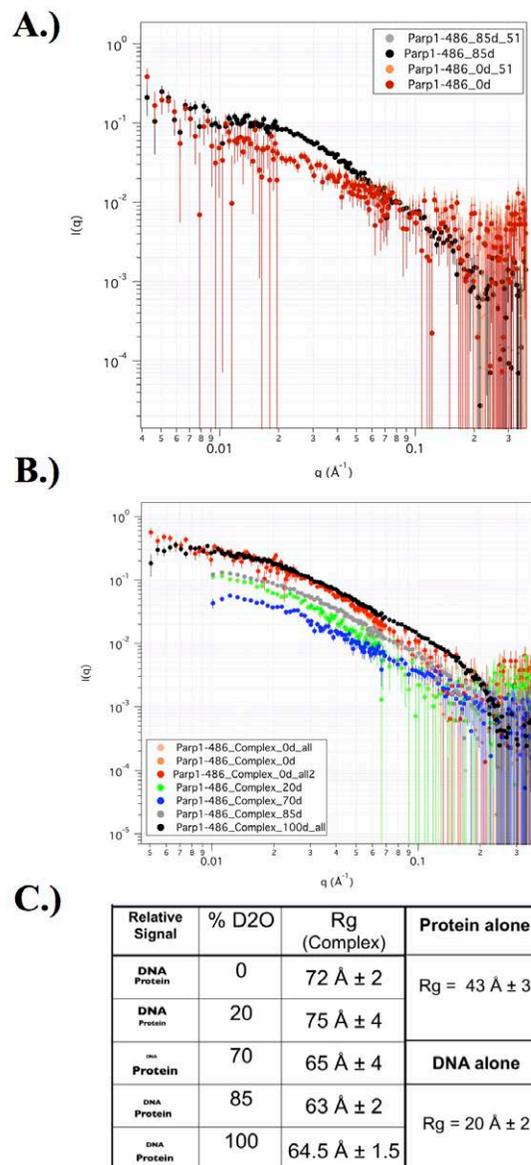


Figure 5.9. SANS contrast variation data on Parp1-486 protein and its complex with the 21-27bp overhang DNA. **A.)** The scattering profiles of Parp1-486 alone. Only two major contrasts were chosen for Parp1-486, the 0% D₂O condition and the 85% D₂O condition. **B.)** The full contrast series of Parp1-486 in complex with DNA (0%, 20%, 70%, 85% and 100% D₂O). **C.)** A table containing Rg values obtained in the full contrast series of Parp1-486 in complex with an overhang DNA damage model (0%, 20%, 70%, 85% and 100% D₂O). The column containing the “relative signal” is meant to signify the changing contribution of signal intensity from either DNA or protein as the percentage of D₂O increases (top to bottom).

Table 5.3. Molecular parameters derived from Small Angle Neutron and X-ray Scattering for Parp1-486 and its complexes with 21-27 bp overhang DNA. The entire data calculated for the contrast-variation series of Parp1-486 and its DNA complexes. The data for Parp1-486 are slightly better than for the full-length because the truncated protein is more monodisperse. The Parp1-486 datasets are more complete because there was an additional data collection for the Parp1-486.

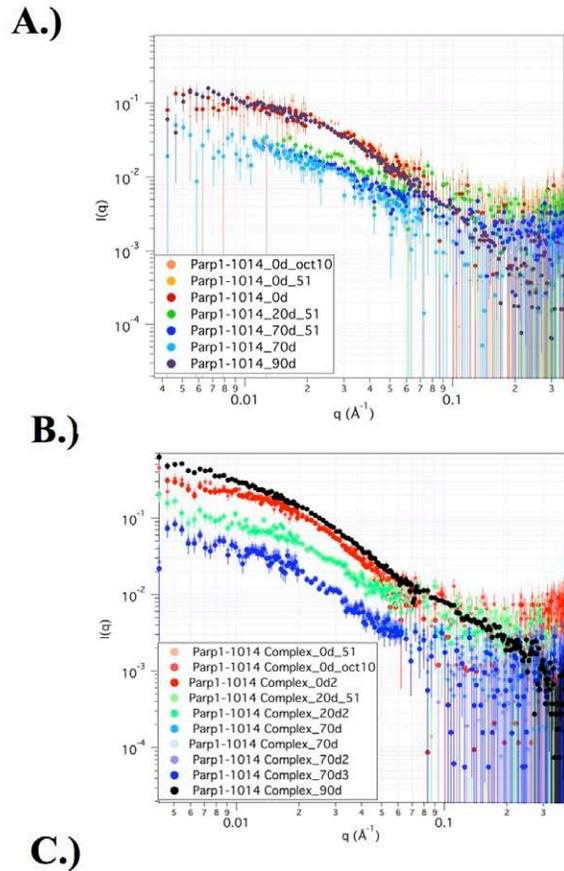
| % D2O | PARP486 | DNA | Complex | Measured Rg (Å) | Total Concentration (mg/ml) | Calculated I(0) (cm ⁻¹) | Measured I(0) (cm ⁻¹) | Calculated Mw (kDA) | Measured Mw (kDA) |
|----------------------|---------|-----|---------|----------------------|-----------------------------|-------------------------------------|-----------------------------------|---------------------|-------------------|
| 0 (6/2010) | X | — | — | 43 ± 3 49 ± 3 | 2.6 | 0.073 | 0.089 0.091 | 55.6 | 68 |
| 0 (6/2010 – lower q) | X | — | — | 53 ± 5 | 2.6 | 0.073 | 0.097 | 55.6 | 74 |
| 0 (10/2010) | X | — | — | 56 ± 4 | 2 | 0.056 | 0.066 | 55.6 | 66 |
| Xray | X | — | — | 44 ± 1 48 ± 1 | 2 | — | — | — | — |
| X-ray | X | — | — | 55 ± 1 56 ± 1 | 1 | — | — | — | — |
| 0 (6/2010) | — | — | X | 72 ± 2 72 ± 5 | 3 | 0.12 | 0.34 0.32 | 70.2 | 200 |
| 0 (10/2010) | — | — | X | 69 ± 5 69 ± 3 | 2 | 0.08 | 0.11 0.1 | 70.2 | 97 |
| 20 (6/2010) | — | — | X | 75 ± 4 72 ± 5 | 3.1 | 0.04 | 0.136 0.127 | 70.4 | 240 |
| 70 (6/2010) | — | — | X | 65 ± 4 65 ± 1 | 2.9 | 0.032 | 0.067 0.064 | 70.8 | 148 |
| 70 (10/2010) | — | — | X | 70 ± 10 73 ± 3 | 1.3 | 0.014 | ~0.016 0.011 | 70.8 | ~81 (large error) |
| 85 (6/2010) | — | — | X | 63 ± 2 | 2.8 | 0.081 | 0.149 | 71 | 130 |
| 100 (6/2010) | — | — | X | 64.5 ± 1.5 | 2.8 | 0.157 | 0.315 | 71.2 | 141 |
| Xray | — | — | X | 62 ± 1 66 ± 3 | 1.3 | — | — | — | — |
| Xray | — | X | — | 24 ± 1 24.5 ± 0.5 | 2 | — | — | — | — |

within the SANS measurements time restraints. Concentrated Parp1-486 protein (~6.0 mg/ml) and 21-27mer DNA (~20.0 mg/ml) were dialyzed in ~100% D₂O complex buffer (300 mM NaCl, 50 mM Tris pH=8). Dialysis of the protein alone resulted in ~40% loss of the initial material due to sticking to the dialysis membrane (loss detected by absorbance readings at 280nm before and after dialysis). Because sticking is a form of aggregation, extra precautions to prevent aggregation during SANS data collection were necessarily applied. It is known that small molecule additives, or excipients, can increase solubility of properly folded proteins and stabilize unfolded proteins [246]. Addition of L-arginine at a 5 mM concentration was sufficient to prevent sticking/aggregation of both Parp1-1014 and Parp1-486 in dialysis, when D₂O was present. SEC-MALS experiments showed that the complexes were stable and behaved nearly identically to conditions with and without L-arginine (same elution times and peak widths as seen in **Figure 5.4**). However, having L-arginine present during the process of concentrating the protein to ~6mg/ml caused the Parp1-1014 protein to undergo a slight self-association in the form of a dimeric species (a 2-6% fraction of a 200kDa species from SEC-MALS). The self-association seen most likely was an interaction mediated by two potential surface exposed cysteine side-chains (likely, the very same cysteines used for maleimide-linkage of fluorophore, **Chapter 2**). Alternatively, the dimerization could result from promiscuity of any number of the Zinc-coordinating cysteine residues found in Zn1-Zn3. Disulfide bonds that disrupt the coordination of zinc, when reversed might render the participants unfolded, and more likely to aggregate, since it is unlikely the Zn would return to its properly bound state even after the cysteines were reduced. Regardless, the unfortunate appearance of a dimeric species was especially noticeable as a left-hand shoulder in the elution peak of the Parp1-1014 and its DNA-complex (**Figure 5.4B**). Past experiments have shown that a reversal of a dimer in SDS-PAGE could be easily achieved through the addition of reducing agent however, depending on the cysteine involved, reducing the disulfide linkages may not resuscitate the structure of a domain if the loss of Zn-coordination resulted in a loss of the Zn-atom in that domain (data not shown). We therefore added TCEP (*tris*(2-carboxyethyl)phosphine) to a final concentration of 5mM, a concentration in nearly 1000-fold excess of the protein. We performed the experiment in the presence of TCEP for the duration of three days.

SANS measurements for the Parp1-486 and its complex with the 21-27mer 3'-overhang DNA were performed in 0%, 20%, 70%, 85% and 100% D₂O at concentrations between 3.1 and 1.3mg/ml (SANS parameters for the Parp1-486 and its complex with DNA are tabulated in **Table 5.3**). The SANS profiles for the complete contrast series are overlaid in for the Parp1-486 protein and the Parp1-486:DNA complex in **Figure 5.9A** and **Figure 5.9B**, respectively. Overall the SANS results matched the SAXS results in terms of R_g changes in the presence of DNA (**Table 5.3**). Most telling was the result of the complex in 70% D₂O contrast variation condition. For Parp1-486 in the 70% D₂O condition, the match point of the DNA, indicates the R_g of protein alone was about 20 Å greater than the protein at the same condition without DNA bound. R_g values for the proteins and the protein-DNA complexes in the different D₂O conditions are represented in (**Table 5.3**).

SANS analysis of Parp1-1014 and its DNA complexes

The Parp1-1014 and the Parp1-1014:DNA complex contrast series includes a the following D₂O percentages; 0%, 20%, 70% and 100%. The resulting SANS scattering profiles for the contrast series for the Parp1-1014 and the Parp1-104:DNA complexes are presented (**Figure 5.10A** and **Figure 5.10B**, respectively). From **Table 5**, some of the $I(0)$ -values for the different contrast conditions were higher than expected (especially as the percentage of D₂O increased above 20%), but the alternate method of determining molecular mass, in addition to the molecular masses obtained in SEC-MALS seems to strengthen the data (**Table 5.2** and **Table 5.5**). Along with the R_g values determined from the low- q region of the scattering profiles of the molecules, a second parameter can be extracted from the same region, the intensity values at the a scattering angle of zero, $I(0)$. $I(0)$ -values can be obtained by extrapolation and are used to calculate the molecular mass of a molecule (reference). Traditionally, the molar mass value is one of the strongest controls for well-behaved systems in small angle scattering studies [247]. From **Table 5.3** and **Table 5.4** the $I(0)$ -calculations of many of the samples had very high values close to 2-fold higher than the expected molar mass of the complexes. At a glance we see that a



C.)

| Relative Signal | % D2O | Rg (Complex) | Protein alone |
|-----------------|-------|--------------|---------------|
| DNA Protein | 0 | 79 Å ± 2 | Rg = 56 Å ± 6 |
| DNA Protein | 20 | 82 Å ± 6 | |
| DNA Protein | 70 | 80 Å ± 7 | DNA alone |
| DNA Protein | 90 | 83 Å ± 6 | Rg = 20 Å ± 2 |

Figure 5.10. SANS contrast variation data on the Parp1-1014 protein and its complex with the 21-27bp overhang DNA. **A.)** The scattering profiles of the Parp1-1014 alone. At conditions of 0% D₂O, 20% and 70% D₂O the full-length enzyme was reasonably “well-behaved” and the R_g-values were very close to those seen by SAXS. The highest concentrations of D₂O were problematic, in terms of aggregation, but the tendency to aggregate seemed to be somewhat dependent on the protein concentration as well. The scattering profiles were very similar to those seen in SAXS. **B.)** The contrast series of Parp1-1014 in complex with DNA (0%, 20%, 70%, and 90% D₂O). **C.)** As in Figure 9 this table contains the crucial parameters obtained in the contrast series of Parp1-1014 in complex with an overhang DNA damage model (0%, 20%, 70% and 90% D₂O). The column containing the “relative signal” signifies the changing contribution of signal intensity from either DNA or protein as the percentage of D₂O increases (top to bottom).

Table 5.4. Molecular parameters derived from Small Angle Neutron and X-ray Scattering for full-length PARP-1 and its complexes with 21-27 bp overhang DNA. The measured I(0) value was used for calculating molar mass of the complex (reference). Under unfavorable experimental conditions, the measured molar mass will be larger than predicted. High D₂O conditions may contribute to protein aggregation because D₂O is known to reduce the solubility of protein (reference). Thus, low percentage D₂O solutions should provide more reliable information. An alternative method to calculate molar mass of molecules in SAS is shown in Table 5. As an aside, the Parp1-1014-DNA complex in 70% D₂O was well behaved on multiple occasions. The calculated and measured values for the protein-DNA complex are in very good agreement, which adds to the confidence in this crucial sample condition.

| % D ₂ O | FL PARP | DNA | Complex | Measured Rg (Å) | Total Concentration (mg/ml) | Calculated I(0) (cm ⁻¹) | Measured I(0) (cm ⁻¹) | Calculated Mw (kDA) | Measured Mw (kDA) |
|-----------------------|---------|-----|---------|----------------------------------|-----------------------------|-------------------------------------|-----------------------------------|---------------------|-------------------|
| 0 | X | — | — | 54 ± 6 | 2 | 0.12 | 0.1 | 114 | 95 |
| 85 | X | — | — | 76 ± 1 | 3.4 | 0.21 | 0.32 | 116 | 180 |
| 90 | X | — | — | 72 ± 2 | 1.2 | 0.092 | 0.12 | 116 | 150 |
| Xray (0) | X | — | — | 60 ± 1 | 2 | — | — | — | — |
| Xray (0) | X | — | — | 60 ± 1 | 1 | — | — | — | — |
| Xray (100) (higher q) | X | — | — | 62 ± 1 | 0.6 | — | — | — | — |
| 0 | — | — | X | 79 ± 2 | 2.2 | 0.15 | 0.27 | 129 | 230 |
| 20 | — | — | X | 82 ± 6 | 2 | 0.042 | 0.076 | 130 | 235 |
| 70 | — | — | X | 80 ± 7 | 1.3 | 0.032 | 0.029 | 131 | 120 |
| 90 | — | — | X | 85 ± 1 (poor fit, aggregates) | 2.2 | 0.17 | 0.37 | 131 | 285 |
| 90 (lower q) | — | — | X | 104 ± 2 | 2.2 | 0.17 | 0.45 | 131 | 350 |
| Xray (0) (higher q) | — | — | X | 73 ± 1 | 1 | — | — | — | — |
| Xray (100) (higher q) | — | — | X | 84 ± 1 | 2 | — | — | — | — |

large proportion of the samples that have larger than expected $I(0)$ -values. Heterogeneity in the system (like a small proportion of dimeric PARP-1 from **Figure 5.4B**), slight differences between the sample buffer (used for buffer signal subtraction) and addition of oligonucleotides, and poor determination of initial protein concentrations can lead to sample irregularities [247]. Additionally, prior to incorporating the use of L-arginine into the sample buffer, a large amount of what appeared to be protein remained bound to the walls of the SANS sample cuvetts. The apparent sticking seemed to be alleviated in future experiments when L-arginine was present but the phenomenon may contribute to the $I(0)$ -derived molar mass inconsistencies.

Upon close inspection of the Parp1-1014 and the Parp1-486 data tables (**Table 5.3** and **5.4**) it is apparent that overall, the lower concentration samples behave better than the higher concentration samples (i.e. less polydispersity manifesting itself as error in the Guinier region, the region where R_g values are calculated). The trade-off to having lower concentration samples is really evident in that those samples have lower data intensity and worse signal-to-noise ratios. To achieve the best SANS data for the full-length PARP-1-DNA complexes, in the lower concentrations (~ 1.0 mg/ml), especially at the critical DNA match point condition of 70% D_2O , these samples were measured in the SANS instrument for extended periods to increase signal to reduce noise in the data set. Because there was a large uncertainty in the form of large error bars and the loss of a well defined plateau region in the low- q , the SANS data of Parp1-1014 in the higher percentages of D_2O , $\geq 70\%$, was not used due to the tendency to aggregate in the experimental concentrations that were used (noticeably larger R_g -values and associated errors). Interestingly, SAXS data collected in 100% D_2O did not display the same trend.

SANS studies on the complexes of the full-length PARP-1 with the 21-27mer overhang DNA reveal a large positive change in R_g , from 54 to 80 Å compared to the un-complexed Parp1-1014 (**Table 5.4**). The $I(0)$ -value for the Parp1-1014 in complex with the 21-27mer 3'-overhang DNA at that 70% D_2O condition indicates a molecular mass of ~ 120 kDa, a value within experimental error of the expected

Table 5.5. Porod volume calculations for the SAS profiles of the PARP-1 proteins and PARP1-DNA complexes. Porod volumes were calculated from the Porod region of the various scattering profiles according to Fisher *et. al.* The method approximates the molar mass of the samples by assuming an average density of 1.37g/cm³. For the DNA-protein complexes the density will not match precisely, however, assuming that the complex is in a 1:1 ratio the DNA would only comprise roughly 11% of the mass. The molecular masses calculated from the Porod volumes are close to the values expected for the samples. Porod volumes might be a reasonable alternative to using the I(0)-calculations since the very low-q regions of many of the samples might contain scattering signal from the contaminating oligomeric species (Calculations were performed on the web server <http://www.if.sc.usp.br/~saxs/>).

| Molecule | Estimated Porod Volume (Å ³) | MW from Porod Volume kDa (Relative Discrepancy) | MW kDa (Calculated) | Rg from Pair Distribution (Å) |
|--|--|---|---------------------|-------------------------------|
| Parp1-486 (0% D2O) | 132,372 | 70.1 (±24.1%) | 56.5 | 43.1 ± 3.1 |
| Parp1-486 (0% D2O) SAXS | 97,794 | 58.6 (±4.6%) | 56.5 | 45.7 ± 0.2 |
| Parp1-486 (85% D2O) | 139,495 | 74.2 (±31.3%) | 56.5 | 50.7 ± 1.4 |
| Parp1-486 (100% D2O) SAXS | 91,549 | 61.8 (±10.4%) | 56.5 | 46.0 ± 0.2 |
| Parp1-486 DNA Complex (0% D2O) | 381,740 | 211.7 (±67.4%) | 71.3 | 68.9 ± 0.8 |
| Parp1-486 DNA Complex (20% D2O) | 350,643 | 194.0 (±63.0%) | 71.3 | 68.9 ± 2.0 |
| Parp1-486DNA Complex (70% D2O) | 140,039 | 74.5 (±24.0%) | 71.3 | 68.7 ± 2.5 |
| Parp1-1014 (0% D2O) | 191,565 | 91.2 (±19.3%) | 113.1 | 59.9 ± 2.3 |
| Parp1-1014 (0% D2O) SAXS | 203,697 | 110.6 (±2.2%) | 113.1 | 60.7 ± 2.0 |
| Parp1-1014 (20% D2O) | 213,808 | 109.6 (±3.1%) | 113.1 | 61.4 ± 0.3 |
| Parp1-1014 (70% D2O) | 182,949 | 79.2 (±29.9%) | 113.1 | 63.8 ± 3.0 |
| Parp1-1014 (100% D2O) SAXS | 201,723 | 109.5 (±3.1%) | 113.1 | 57.9 ± 0.4 |
| Parp1-1014 DNA Complex (0% D2O) SAXS | 259,059 | 180.7 (±39.0%) | 127.8 | 73.1 ± 0.9 |
| Parp1-1014 DNA Complex (0% D2O) | 303,514 | 167.3 (±28.7%) | 127.8 | 79.2 ± 1.7 |
| Parp1-1014 DNA Complex (20% D2O) | 252,167 | 122.0 (±3.1%) | 127.8 | 83.9 ± 7.6 |
| Parp1-1014 DNA Complex (70% D2O) | 214,054 | 116.5 (±3.1%) | 127.8 | 82.4 ± 2.2 |
| Parp1-1014 DNA Complex (100% D2O) SAXS | 276,900 | 193.3 (±48.7%) | 127.8 | 73.0 ± 0.3 |
| Parp487-1014 (0% D2O) SAXS | 143,950 | 76.7 (±26.0%) | 56.8 | 39.7 ± 0.1 |

value of 131 kDa, a value that is corroborated using the Porod volume to calculate molecular mass (**Table 5.5**). Of note, a ~4kDa increase in molecular mass for the Parp1-1014:DNA complex should be expected as a result of Hydrogen/Deuterium exchange in the higher percentages of D₂O, hence the 131 kDa expected molecular mass for the 70% D₂O condition (larger than the theoretical mw of 127.8kDa in H₂O). Additionally, the R_g-measurement indicates that the protein at the DNA match point is itself 20Å larger than the protein in the unbound form. That result is indicative of a scenario in which the DNA in the complex is not contributing much to the length of the complex. Moreover, the R_g-value for the protein in the 20% D₂O condition, the condition where the DNA signal is stronger than the protein signal, is not greatly different than the 70% D₂O condition. Furthermore, the R_g-values of the proteins and the protein-DNA complexes obtained by SAXS are in close agreement with the SANS data. Additional evidence of a conformational change is available in the change in Stokes' radii (R_s) for the protein-DNA complexes since the elution profiles were compared with column calibration standards, and Stokes' radii calculated for each. Furthermore, the change in Stokes's radii, or elution volume, also reflects a molecular mass that is appropriate for a 1:1 PARP:DNA complex as seen in SEC-MALS data (as seen in **Figure 5.4** and tabulated in **Tables 5.1** and **5.2**).

Parp1-1014 is more susceptible to protease digestion upon binding DNA.

In order to confirm that indeed PARP1 undergoes a conformational change upon binding DNA, another limited proteolysis experiment was performed. A careful search for alternative proteases revealed that another serine protease family member, thrombin, is expected to cleave PARP1 in just one site (residue 496, prediction with the aid of http://web.expasy.org/peptide_cutter/).

As a means to prevent DNA from occluding any possible protease cut-sites, a smaller 14-17mer 3bp 3'-overhang DNA was used. Because PARP1 is thought to protect ~14bp of DNA (7bp for each of the Zn-fingers in the N-terminal DNA binding region), a DNA of approximately 14bp should be sufficient to reproduce the conformational change but not too large to prevent the protease from reaching

any potential cut-sites near the vicinity of the DNA. To prevent any possible molecular crowding events that might enhance protease activity DNA was titrated into reactions on a fractional basis. After digesting the Parp1-1014 protein with increasing units of thrombin, an optimal value of 0.5 units was chosen for DNA titration experiments (**Figure 5.8A**). Next, Parp1-1014 was incubated with increasing amounts of DNA for 30 minutes prior to adding thrombin. From those thrombin digestions it became obvious that Parp1-1014 was indeed more sensitive to protease digestion with DNA present (**Figure 5.8B**). Additionally, when DNA was present the digestion patterns appeared markedly different than when it was absent. These results suggest that DNA imparts a change on the protein and may in fact indicate that PARP1 maintains a more open structure in the presence of DNA.

Ensemble atomistic models reveal the probable regions of PARP-1 that are responsible for the increased length seen when it complexes with DNA damage models.

Small angle scattering data collected on biological macromolecules can yield low-resolution structural information. Scattering data are usually displayed as a plot of $I(q)$ vs. q , where I is the intensity vs. q the momentum transfer (**Figure 5.5**). It is often taken for granted that 3-D structural information, albeit low-resolution, can be calculated from such a curve. *Ab initio* envelopes can be calculated using spherical harmonics but these envelopes do little to help one envision how certain structural regions are positioned in solution (see **Chapters 3 and 4**). To better understand the low-resolution structure of a multi-domain protein, the positions of structural features like the positions of individual domains can be modeled and these models judged by comparison of calculated SAXS profiles to experimental data. To gain structural insights on how a molecule behaves in solution, small angle scattering data can be used in conjunction with high-resolution crystallography or NMR data. Since so much structural information is known for PARP1, it is a great candidate for molecular dynamics simulations. Many computer programs have been designed to help build and refine atomistic models. Since solution studies are low-resolution in nature, at best only a family of models that fit the experimental data can be obtained. Atomistic modeling

programs generate large numbers of possible structures (up to 10^5), known as ensembles. Ensembles are thought to represent the shape heterogeneity of the molecules in question by comparing results to the low-resolution SAS data, which is representative of the bulk average scattering profile of the solution. The most effective programs allow for a comparison of modeled data to experimental data, so one can evaluate the goodness-of-fit for the model ([232, 248]). Some programs generate theoretical scattering profiles for just such comparisons to experimental data. One such program suite, SASSIE has recently been developed to model structures of intrinsically disordered proteins and their protein or nucleic acid complexes [232].

The combined structural data from X-ray crystallographic and NMR studies allowed us to create atomistic models for Parp1-486, Parp487-1014 and Parp1-1014 (see methods). Starting structures for the proteins alone were in great agreement with the data since the predicted scattering profiles for the proteins had χ^2 -values ~ 1 (data not shown). The full-length PARP-1 model was further explored to determine if adding DNA on the very edge of the N-terminus would result in a ~ 30 Å increase in R_g . Models for three such DNA positions are depicted in **Supplementary Figure 5.3**. The results show that even if DNA was bound on the very N-terminal end of the protein, a highly unlikely position considering that the Zn2 domain has a higher affinity for DNA than does the Zn1 domain, the R_g -values for the simulated model are far less than what is seen experimentally (**Table 5.6**). As a means to investigate likely domain-domain interaction regions, the surface electrostatics of one of the full-length PARP-1 atomistic structures were investigated using a modified version of the Poisson-Boltzman equation [243]. No obvious charged patches or regions where domains might interact were immediately apparent (**Supplementary Figure 5.4**). Once the starting structures were built and energy minimized we began our hybrid Monte Carlo-Molecular Dynamics simulations in the SASSIE program. In SASSIE, each of the starting structures were centered and aligned and theoretical scattering profiles were calculated. Next, the boundaries of the known domains were defined so that only the linker regions were allowed to vary in the structure-generating program. The linker boundaries of PARP-1 used in the modeling were between the following pairs of amino acids; linker1 91-112, linker2 202-224, linker3 350-387, linker4 487-528, linker5 640-664

also, the first 5 N-terminal amino acids were allowed to rotate. For our experiments, sets of dihedral angles (ϕ, ψ) for each of the amino acids in each of the linkers were allowed to rotate a maximum of 30° per round. Every round of MMC calculations resulted in a generated structure that was evaluated energetically by the confines defined by the most current molecular dynamics techniques [232]. The spectra from the large ensembles of structures were then filtered according to closeness of fit to experimental scattering profiles.

Multiple runs that generated ensembles of 10^4 structures were performed for each of the molecules studied by SAS. A theoretical scattering profile was then generated for each of the different conformations of the protein or protein-DNA complex. Approximate R_g -values for the ensemble were then plotted against structure number as a means to see the breadth of conformational space each MD run covered (Parp1-486, Parp1-1014, Parp1-486-DNA complex, and the Parp1-1014-DNA complex; **Figure 5.11A**, **Figure 5.12A**, **Figure 5.13A** and **Figure 5.14A**, respectively). The structures were then filtered for goodness of fit to experimental scattering profile for each molecule and the χ^2 -value for the best-fit structures were plotted against R_g (**Figure 5.11-14B** and **Figures 5.11-14C**). From **Figure 5.11** and **Figure 5.12** it is obvious that both the Parp1-1014 and the Parp1-486 molecules are very flexible and that many possible structures fit the data with relatively low χ^2 -values. Also, noticeable is that the structures with the very small and the very large R_g -values also have the highest χ^2 -values. The same is true for the Parp1-486-DNA complex and the Parp1-1014-DNA complex as seen in **Figure 5.13** and **Figure 5.14**, respectively.

To better understand the importance of the ensembles and the solution studies themselves, we incorporated the use of density maps as a means to visualize the extreme flexibility of the molecules. Because the ensembles are atomistic in nature, and are saved in coordinate file format, we can plot the relative position of a given domain or domains for the entire ensemble of 10^4 structures. To further refine our ensembles, we decided to plot density maps of the best-fit structures, which included the structures with the lowest χ^2 -values. For the DNA complexes the χ^2 -values typically ranged from 12 to 1.5. For

best results, we concatenated the structures from 3-4 separate runs of 10^4 structures resulting in an analysis of effectively 30-40,000 structures. The best-fit structures for the proteins alone had low χ^2 -values that ranged from ~ 3 to 0.5. For the isodensity plots for both Parp1-486 and Parp1-1014 we centered and aligned the ensemble around the Zn3-domain (**Figures 5.15** and **5.16**, respectively). For the DNA-bound forms of Parp1-486 and Parp1-1014, the iso-density plots were aligned on the Zn1-Zn2 (DNA binding domain) (**Figures 5.17** and **5.18**, respectively). From **Figure 5.19** one can visualize just how flexible, in terms of the density of the ensembles, and the differences for the Parp1-486 are easily compared to the full-length enzyme.

When all the structures are aligned on the Zn3, both proteins share a common feature the position of the Zn1 and to a lesser extent Zn2 are very dynamic as opposed to the BRCT domain. From the density plots one can see how the Zn1-Zn2 domains appear to be continuously scanning and might explain why the domains are so effective in their ability to find and bind multiple DNA structures. For the structures of the proteins in complex with 3'-overhang DNA, the density maps of the complexes were centered and aligned on the Zn1-Zn2 DNA-bound region. From the density maps of the complexes one can visualize the increase in the extended nature of the proteins once bound to their DNA substrate. The image brings to mind a beads-on-a-string example where the string tethers the bead and only allows a given domain a certain reach depending on its mass and the length of the linker tether. As with the proteins alone, a direct comparison of the ensemble coverage of the proteins in complex with DNA can be seen in **Figure 5.19**. When one compares all four density plots for the Parp1-1014, Parp1-1014:DNA complex, the Parp1-486 and the Parp1-486:DNA complex it becomes very obvious that the ensemble of the truncation construct has a severely diminished maximum distance.

Finally, the atomistic models of the best-fit structures for each run can be examined individually or can be compared to the resulting structures from different runs (**Figure 5.20A**). Figure 20A depicts the resulting best-fit structures from three separate SASSIE runs of the full-length PARP1-DNA complex in the critical 70% D₂O condition. Overall, the structures have very similar χ^2 -values, size dimensions, general shapes and overall very elongated, open appearances. When measurements of the approximate

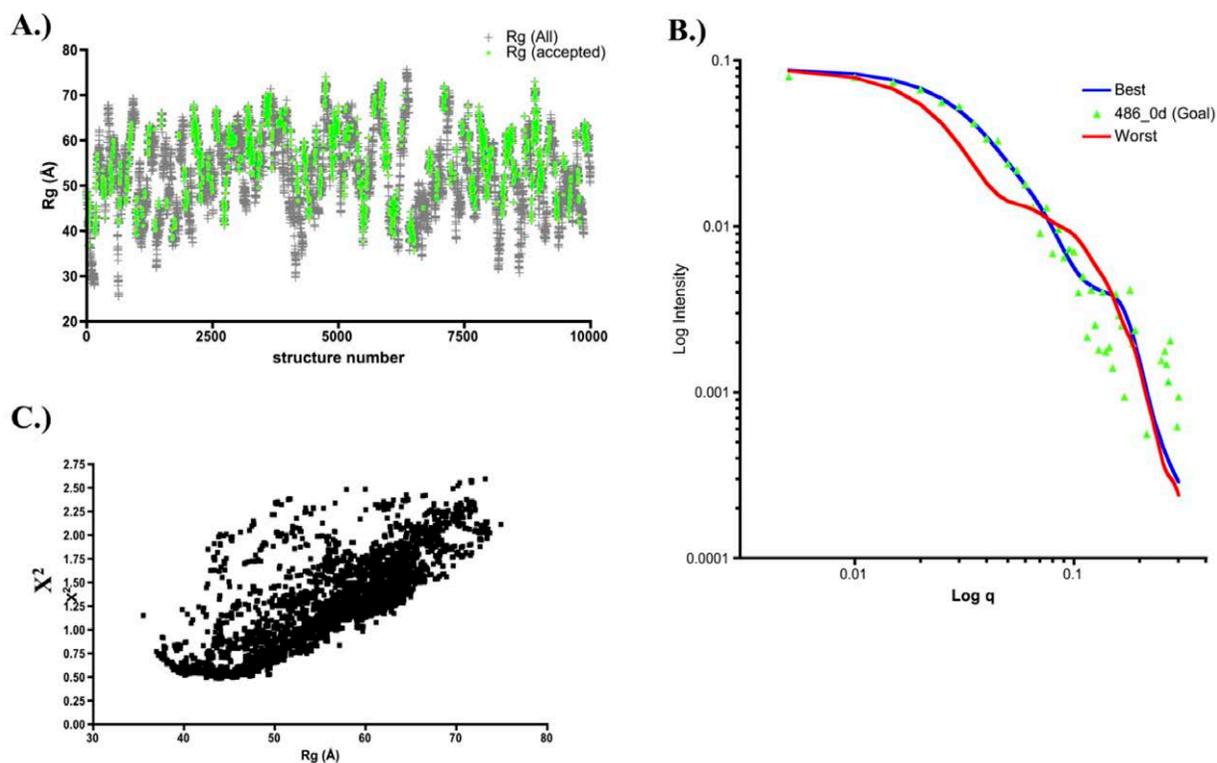


Figure 5.11. Ensembles of possible PARP1-486 conformations present in SAS profiles. **A.)** Using the Configuration Generator module within the program SASSIE, 10^4 possible structures were generated by randomly varying φ and ϕ angles within the linker regions that reside between ordered domains. In black + symbols represent all the configurations that were generated. The green circles represent the accepted structures that were energetically favorable and lacked overlap as confirmed by NAMD minimization using the CHARMM-22 force field. Small Angle Scattering (SAS) profiles were generated for all of the accepted structures using the program CRYSON. From this plot one can see the range of R_g -values possible for the different conformations. **B.)** The scattering profiles for the accepted structures were then filtered according to the fit to experimental SANS data. The plot is of the profiles of the best and worst structures to the PARP1-486 SANS data in 0% D_2O (blue, red and green, respectively). **C.)** A plot of the χ^2 -values vs. the optimal R_g -value for each of the best-fit structures.

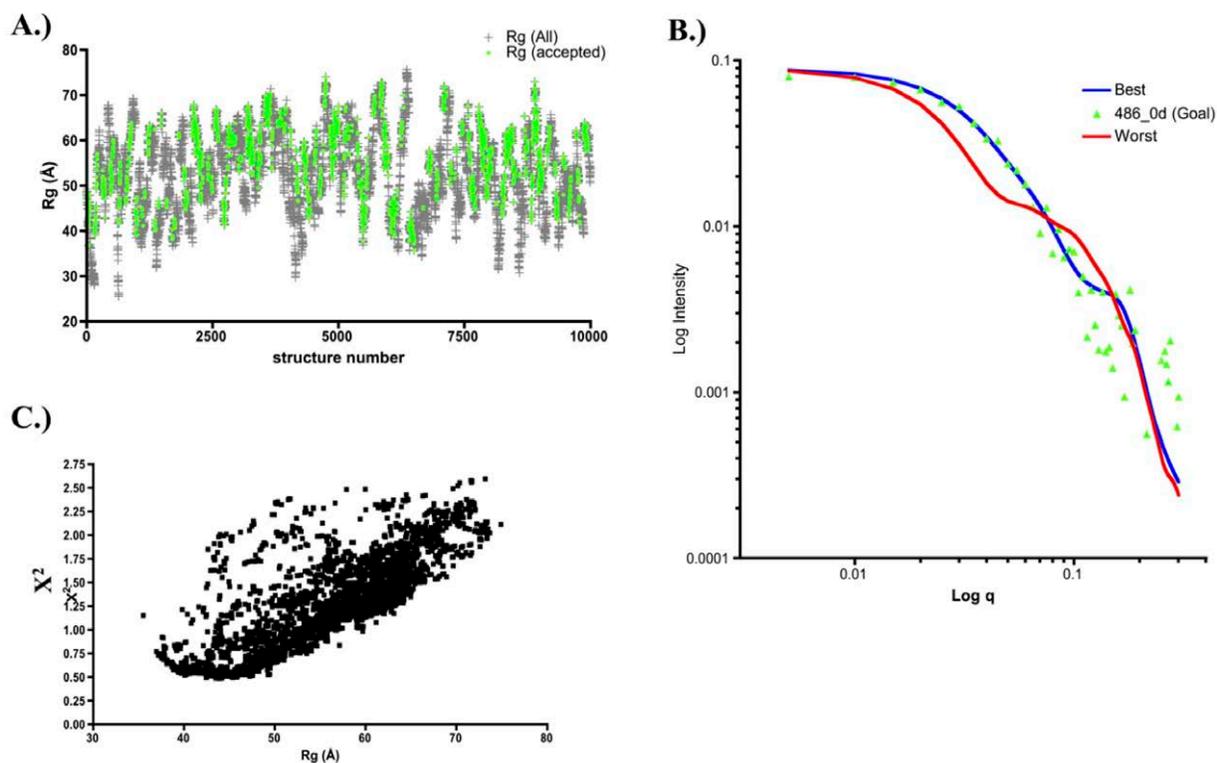


Figure 5.11. Ensembles of possible PARP1-486 conformations present in SAS profiles. **A.)** Using the Configuration Generator module within the program SASSIE, 10^4 possible structures were generated by randomly varying φ and ϕ angles within the linker regions that reside between ordered domains. In black + symbols represent all the configurations that were generated. The green circles represent the accepted structures that were energetically favorable and lacked overlap as confirmed by NAMD minimization using the CHARMM-22 force field. Small Angle Scattering (SAS) profiles were generated for all of the accepted structures using the program CRYSON. From this plot one can see the range of R_g -values possible for the different conformations. **B.)** The scattering profiles for the accepted structures were then filtered according to the fit to experimental SANS data. The plot is of the profiles of the best and worst structures to the PARP1-486 SANS data in 0% D_2O (blue, red and green, respectively). **C.)** A plot of the χ^2 -values vs. the optimal R_g -value for each of the best-fit structures.

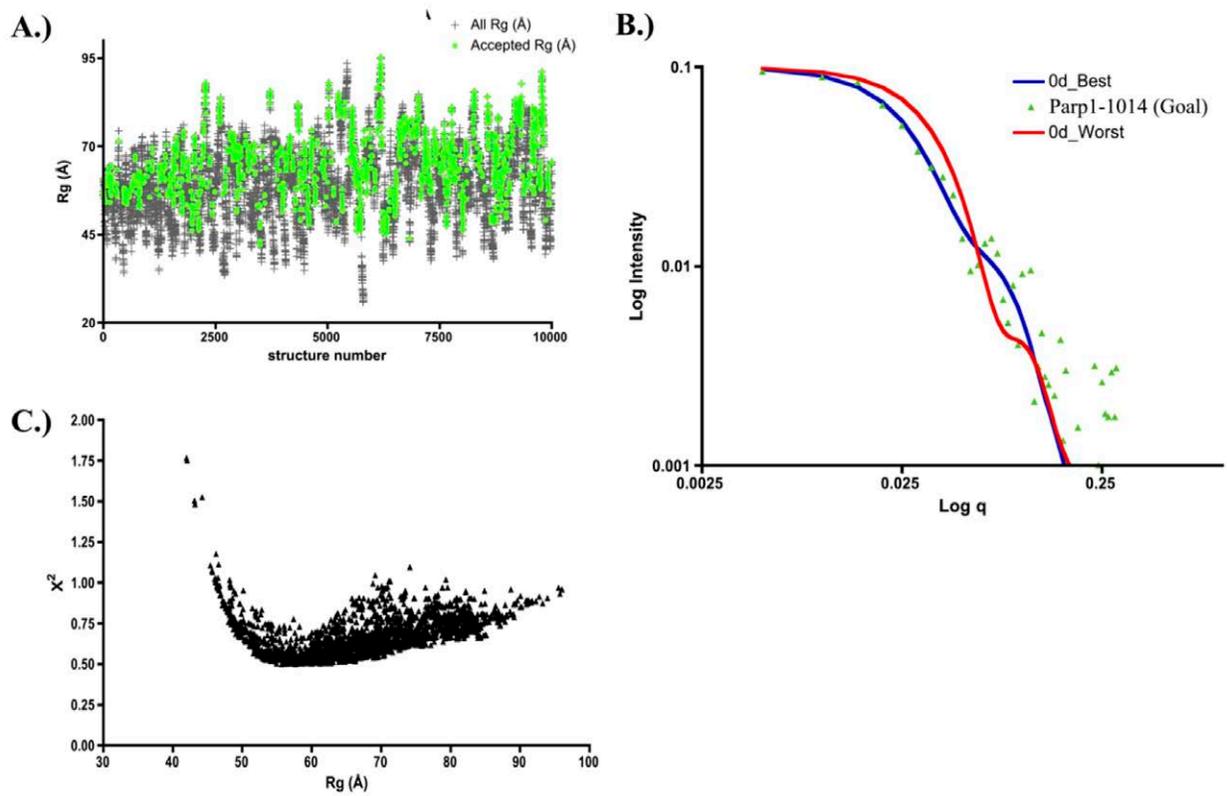


Figure 5.12. Ensembles of possible Parp1-1014 protein conformations present in SAS profiles. Using the same procedure as in Figure 11, models were generated for PARP1-1014, filtered and selected (A), fit to experimental data (B) and χ^2 values plotted (C). The worst χ^2 -values are for the structures with the extreme R_g -values. It is clear that a large number of structures fit the data quite reasonably when following the χ^2 criteria.

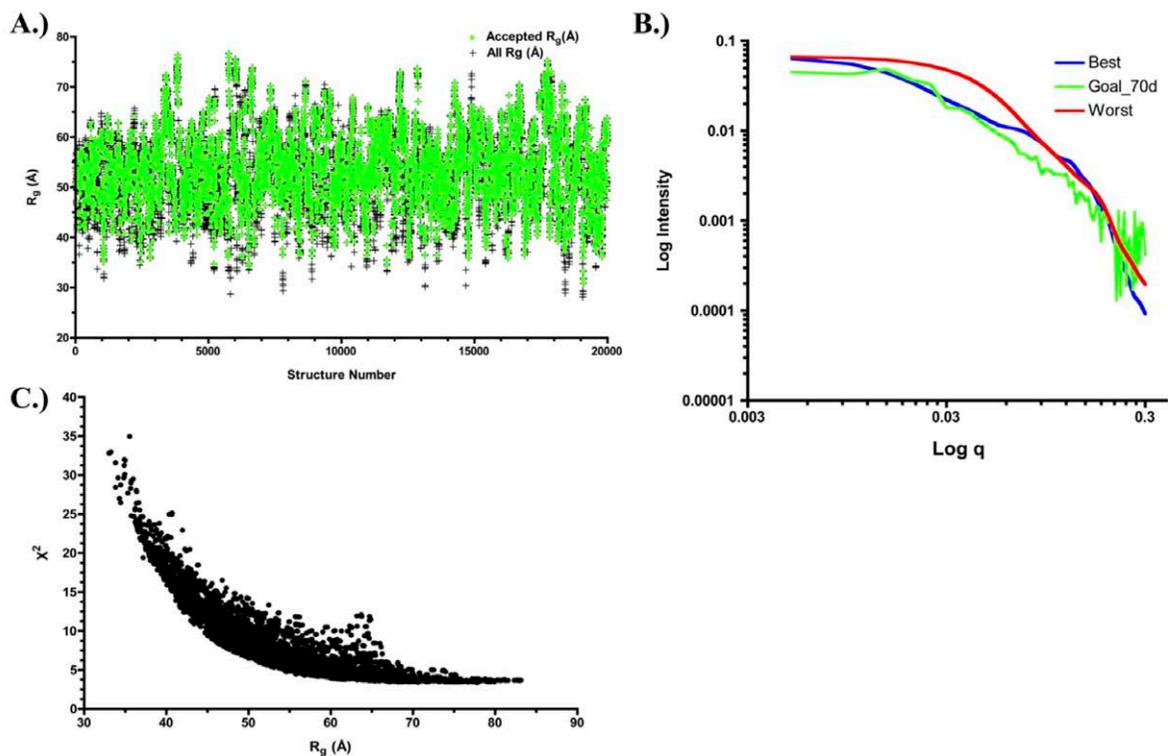


Figure 5.13. Ensembles of possible PARP1-486-DNA complex conformations present in SAS profiles. A) Structures were created and filtered as described in Figure 11. B.) The scattering profiles for the accepted structures were then filtered according to the fit to experimental SANS data. The plot is of the profiles of the best and worst structures to the Parp1-486-DNA complex SANS data in 70% D_2O (blue, red and green, respectively). The best-fit structure had a $\chi^2=3.4$, which is acceptable. C.) A plot of the χ^2 -values vs. the optimal R_g -value for each of the best-fit structures. The worst χ^2 -values are for the structures with the extreme R_g -values. It is clear that a large number of structures fit the data reasonably, however, the greatest density of structures are confined to the R_g -values near 60Å (very close to the R_g -estimate for the protein-DNA complex in 70% D_2O).

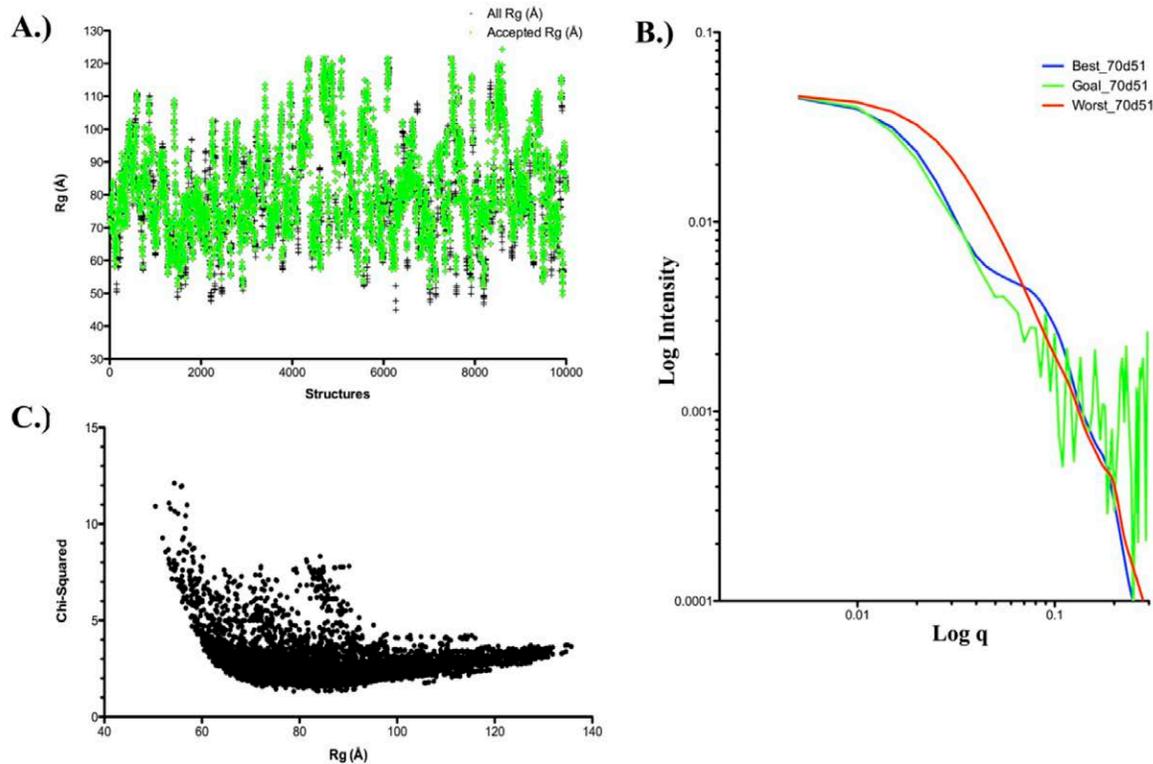


Figure 5.14. Ensembles of possible Parp1-1014-DNA complex conformations present in SAS profiles. **A.)** Structures were generated and filtered as described in Figure 11. **B.)** The scattering profiles for the accepted structures were then filtered according to the fit to experimental SANS data. The plot is of the profiles of the best and worst structures to the Parp1-1014-DNA complex SANS data in 70% D_2O (blue, red and green, respectively). The best-fit structure had a $\chi^2=1.66$ (and without the low- q data the $\chi^2=1.33$). **C.)** A plot of the χ^2 -values vs. the optimal R_g -value for each of the best-fit structures. The worst χ^2 -values are for the structures with the extreme R_g -values. It is clear that a large number of structures fit the data reasonably, however, the greatest density of structures are confined to the R_g -values near 80 Å (very close to the R_g -estimates for the protein-DNA complex in 70% D_2O).

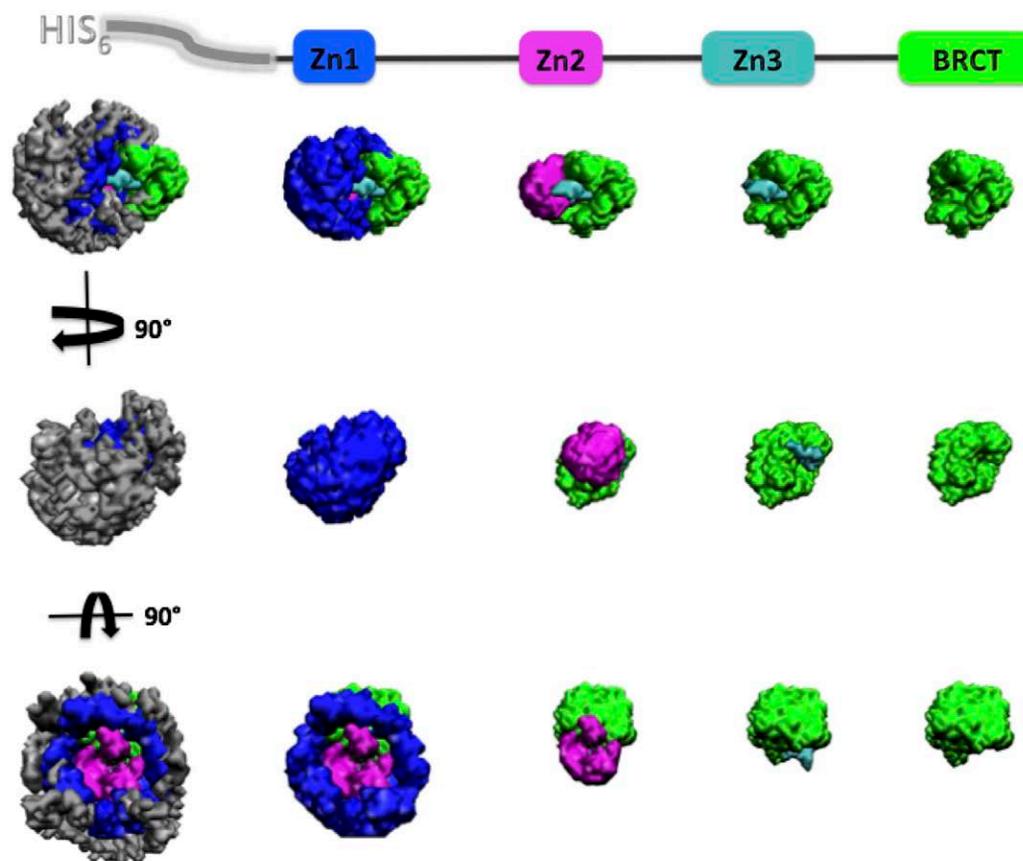


Figure 5.15. Iso-density plots of Parp1-486. The ensemble of accepted structures from several SASSIE, 3 or 4×10^4 , runs were combined and filtered against experimental data. As a means to better visualize the breadth of domain coverage, all of the structures were centered and aligned around the Zn-3 domain. The top cartoon in the figure is a schematic representation of the domain organization of Parp1-486. Under each domain are three different views of the density plot for each domain. The globular nature of each region signifies the area that a particular domain travels within the ensemble of structures.

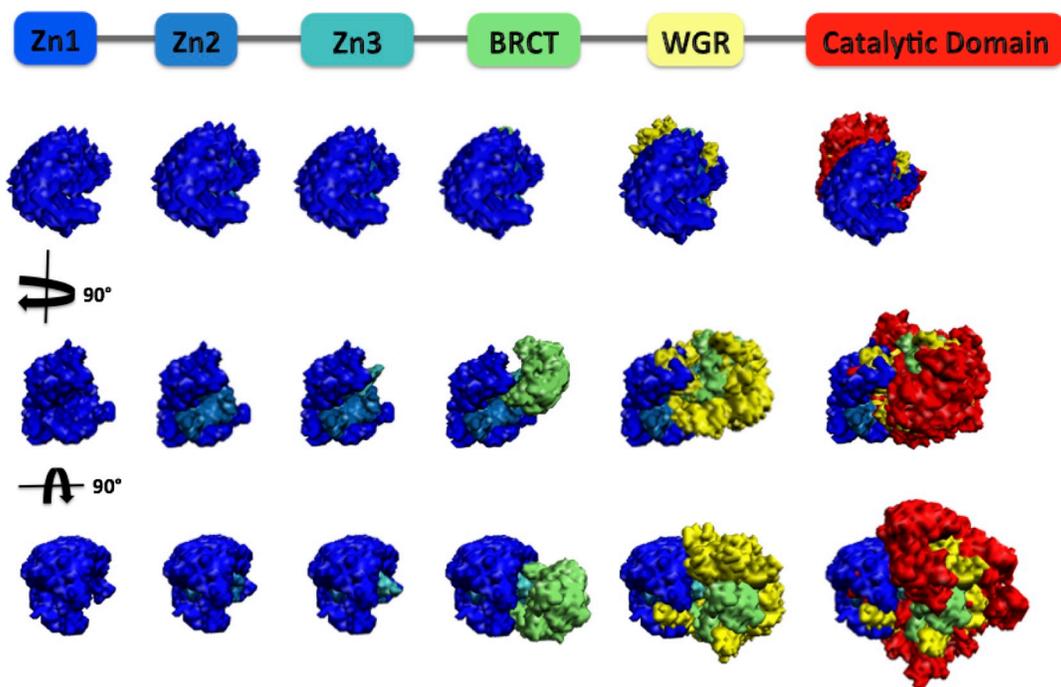


Figure 5.16. Iso-density plots of Parp1-1014. The ensemble of accepted structures from several SASSIE, 3 or 4×10^4 , runs were combined and filtered against experimental data. As a means to better visualize the breadth of distances that each domain can cover, all of the structures were centered and aligned around the Zn-3 domain. The top cartoon in the figure is a schematic representation of the domain organization of Parp1-1014. Just as in the cartoon the domains are colored in a gradient of blue to red and the colors correspond to N to C terminal, respectively. Under each domain are three different views of the density plot for each domain. The globular nature of each region signifies the area that a particular domain travels within the ensemble of structures.

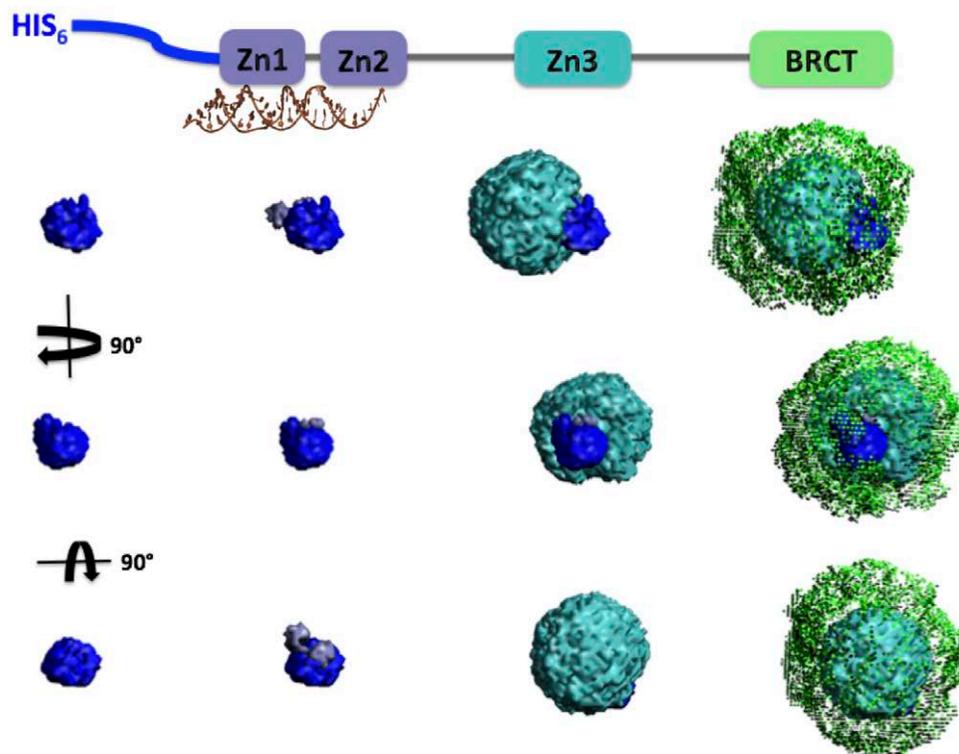


Figure 5.17. Iso-density plots of the Parp1-486-DNA complex. The ensemble of accepted structures from several SASSIE, 3 or 4×10^4 , runs were combined and filtered against experimental data. As a means to better visualize the breadth of distances that each domain can cover, all of the structures were centered and aligned around the Zn1-Zn2 DNA binding region. The top cartoon in the figure is a schematic representation of the domain organization of Parp1-486. Just as in the cartoon the domains are colored in a gradient of blue to red and the colors correspond to N to C terminal, respectively. Under each domain are three different views of the density plot for each domain. The globular nature of each region signifies the area that a particular domain travels within the ensemble of structures.

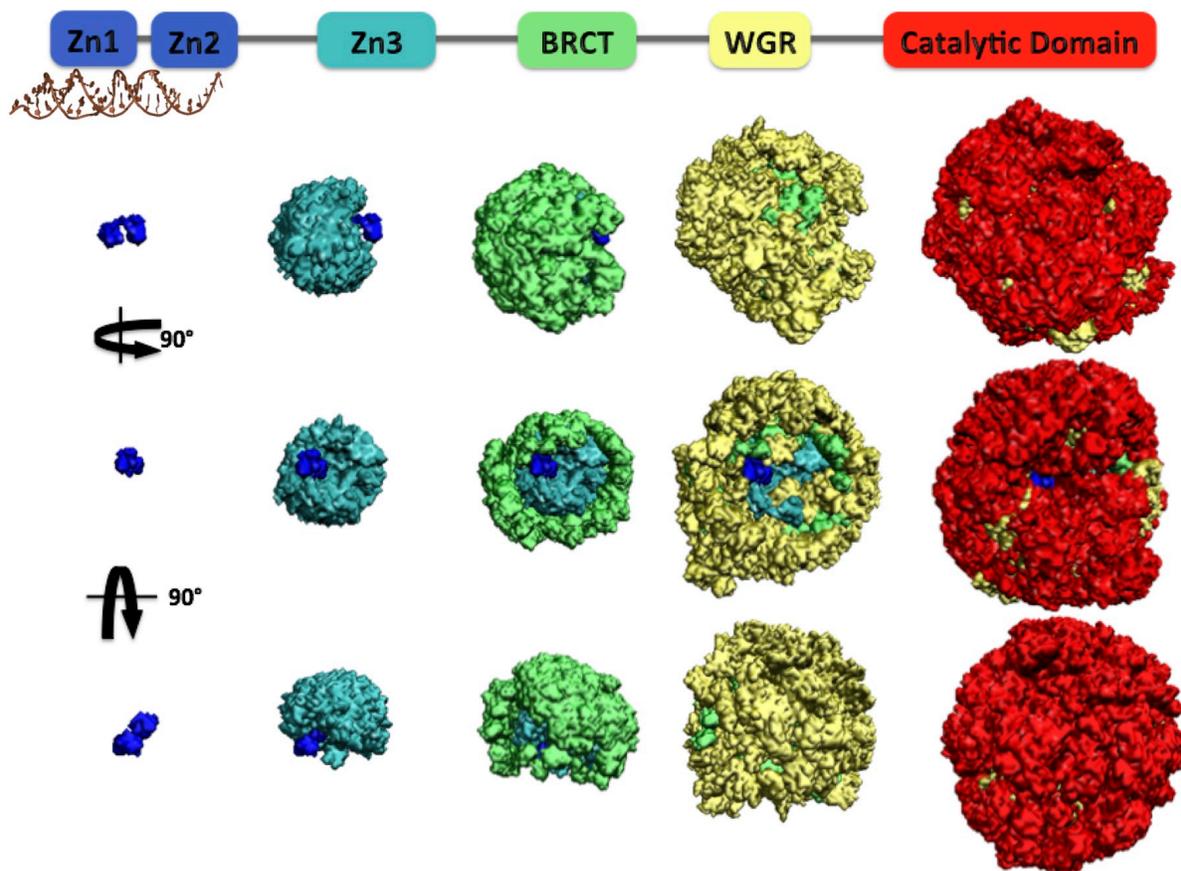


Figure 5.18. Iso-density plots of the Parp1-1014-DNA complex. The ensemble of accepted structures from several SASSIE, 3 or 4×10^4 , runs were combined and filtered against experimental data. As a means to better visualize the distances that each domain can cover, all of the structures were centered and aligned around the Zn1-Zn2 DNA binding region. The top cartoon in the figure is a schematic representation of the domain organization of Parp1-1014. Just as in the cartoon the domains are colored in a gradient of blue to red and the colors correspond to N to C terminal, respectively. Under each domain are three different views of the density plot for each domain (front, side and top). The globular nature of each region signifies the area that a particular domain travels within the ensemble of structures.

maximum distance (D_{\max}) are made in the final structure of the Parp1-1014 or its complex with DNA, the resulting values are very close to the D_{\max} values predicted in the pair distribution plots, $P(r)$, for those molecules (**Figure 5.20B** and **Figure 5.20C**, respectively). For instance the apparent D_{\max} for the Parp1-1014 protein is 183.24Å and for the 21-27mer 3'-overhang DNA:Parp1-1014 complex the apparent D_{\max} was 234Å as compared to the GNOM values of 187Å and 240Å, respectively. Not surprisingly, the linker region containing the thrombin cut-site and separately defined sites of automodification exists as a slightly condensed loop-structure in the Parp1-1014 structure while the same region is fully exposed in one of the best-fit structures of the protein in complex with DNA. Comparing the linker region within the best-fit model Parp1-1014 alone and the model of the protein-DNA complex reveals a large increase in distance the linker spans and is on the order of 37 Å (**Figure 5.20D** and **Figure 5.20E**).

Discussion:

PARP-1 recognizes DNA damage in the cell through its N-terminal DNA binding region, which consists of the Zn1 and Zn2 domains. The recognition of damaged DNA is crucial for the maintenance of the genome [85, 86]. Once PARP-1 finds DNA damage in the form of double stranded breaks, 3'-extensions and nicked regions (referred to in this manuscript as blunt, overhang and nicked damage models) it conveys a signal to the C-terminal catalytic domain, which in turn activates the enzymatic function. The signal that allows for the catalytic domain's activation is not known but was thought to emanate from the hinge-like Zn3 domain [35, 36, 231, 249]. Recently our group described an N-terminal fragment of PARP-1 that was able to undergo a conformational change upon binding DNA damage models. We described an elongation of the Parp1-486 protein in complex with DNA by using SAXS. In this work we attempted to extend the previous study by incorporating Small Angle Neutron Scattering combined with contrast variation and atomistic modeling coupled with low-grain molecular dynamics. Additionally we impart the same studies on the more biologically relevant and functional enzyme, the

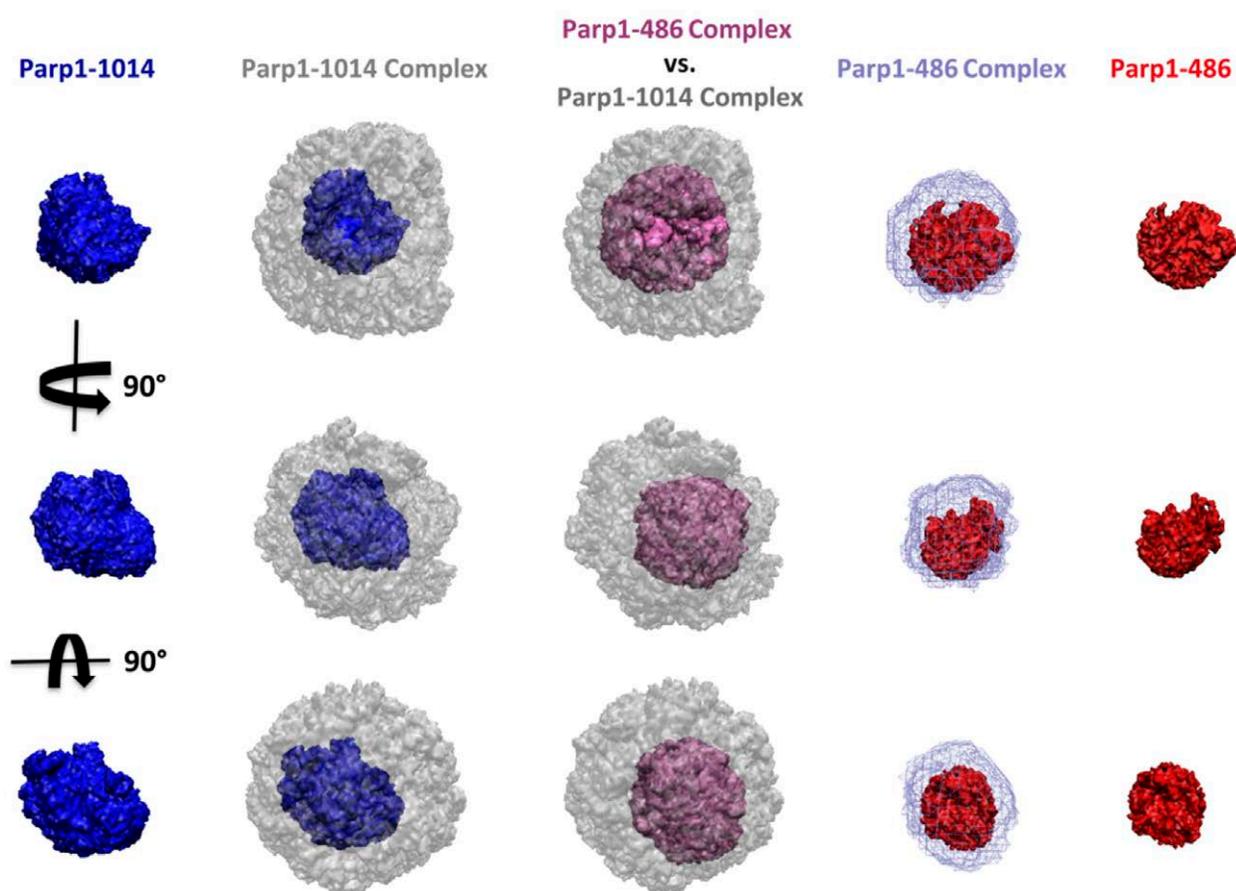


Figure 5.19. A comparison of the different iso-density plots of Parp1-486, Parp1-1014, Parp1-486-DNA complex and the Parp1-1014-DNA complex. As a means to better visualize the differences that each protein or protein-DNA complex can cover, the density plots have been aligned and overlaid. The best-fit density plots for each species gives information on the possible distances the domains can travel within a given protein or protein:DNA complex. When DNA is present in the form of a complex, the density plots grow in volume. The corresponding volume increase when DNA is present and correlates with the increase in R_g -values for the ensembles. As might be expected, the PARP1 proteins have a great deal of flexibility but the greatest density of structures within the χ^2 vs R_g plots of Figures 11-14 reside in the R_g -range defined by the SAS data. The density plots are a good representation of the intrinsic flexibility of the proteins. Even in the absence of DNA the proteins have the potential to expand as much as their linker regions allow, just as they can when DNA is present however, in the absence of DNA, the Parp1-1014 and Parp1-486 are significantly more compact. Filtering the structures against the data reveals the optimal density of the ensemble for each molecule.

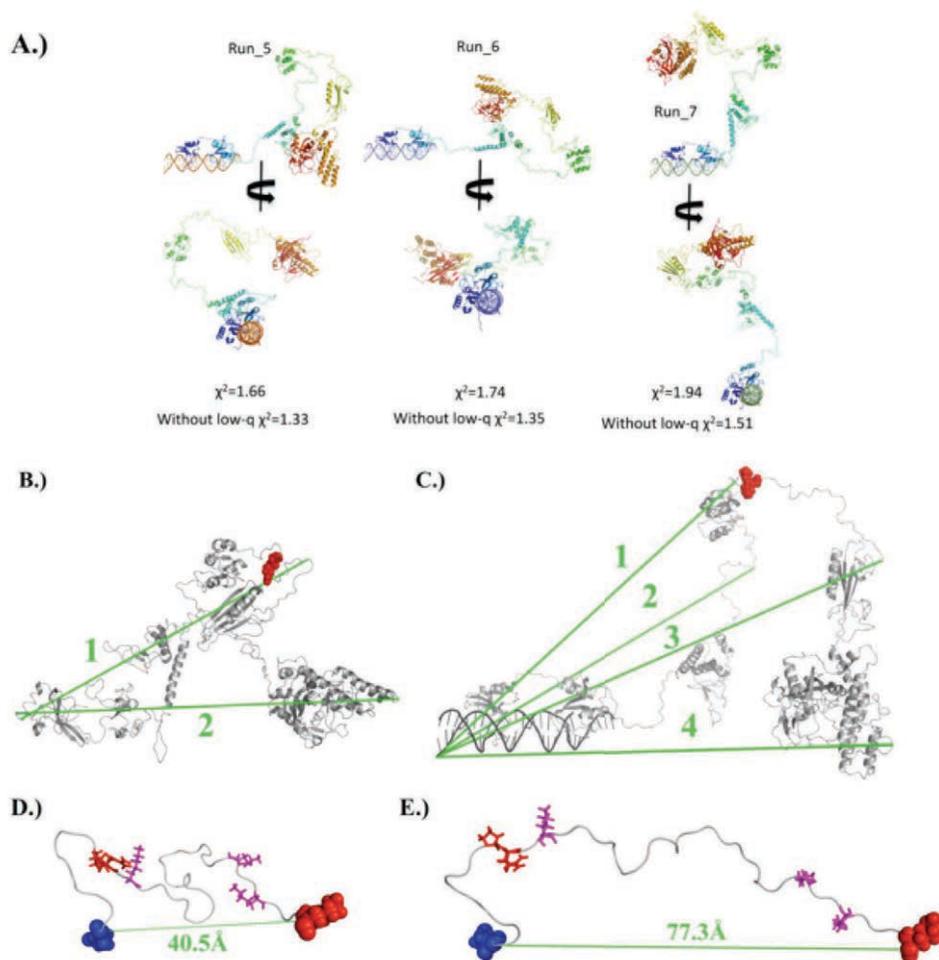


Figure 5.20. The best-fit atomistic structures of the full-length PARP1 enzyme and its DNA complex. **A.)** The best-fit structures of the Parp1-1014 in a complex with the 21-27mer 3'-overhang DNA damage model (head on and rotated 90°). From left to right, the best-fit structures for SASSIE run 5, run 6 and run 7. Two χ^2 -values for each of the structures are listed and χ^2 -values improve further is the low-q data collected at the 13-meter detector distance is not included when filtering the ensemble against the experimental data collected in 70% D₂O, the DNA match point. The structures are colored in a gradient from blue to red from the N- to C-terminus, respectively. **B.)** Measurements of the approximate maximum dimension (D_{\max}) for the best-fit Parp1-1014 structure. Distance 1= 160.83Å and Distance 2=183.24Å.. **C.)** Measurements of the approximate maximum dimension for the Parp1-1014-DNA complex. Distances 1-4 correspond to 197, 204, 234 and 211, respectively. From **B.)** Distance 2 and from **C.)** Distance 3 are the apparent D_{\max} -values for the Parp1-1014 and Parp1-1014-DNA complex, respectively. **D.)** A close-up view of the linker region that resides between the BRCT and WGR domains in the best-fit Parp1-1014 atomistic model. The red sticks comprise the thrombin proteolytic site and are the amino acids proline and arginine. The magenta sticks represent the lysine residues that serve as the major automodification targets (Altmeyer *et al.*, 2009). The blue and red spheres are the N-terminal and C-terminal ends of the linker region, respectively. When comparing the linker region in **E.)**, the best-fit atomistic model for the Parp1-1014-DNA complex, the linker region is considerably expanded in the DNA-complex (40.5Å and 77.3Å, respectively).

full-length PARP-1. The structural information we have obtained is relevant to the DNA repair field because our studies included the full-length PARP-1 in complex with DNA-damage substrates and these complexes were enzymatically active. We have described the DNA-binding induced conformational rearrangements of the full-length enzyme.

Our solution-based studies incorporate several robust methods including small angle x-ray scattering, small angle neutron scattering, size-exclusion chromatography coupled multi-angle light scattering and ensemble structural modeling techniques. The melding of these methods allows for the dynamic description of important yet elusive flexible proteins like PARP-1. When combined with the information discussed in the second chapter of this thesis we have contributed to the PARP1 field by increasing understanding of this protein and its complexes with DNA not only structurally, but by describing its specificity, affinity and enzymatic efficiency in the presence of certain DNA damage models.

Initial SAXS-derived *ab initio* envelopes of the two halves of PARP-1 described very elongated particles (depicted in Chapters 3 and 4). It came as a surprise to see the full-length PARP-1 envelope was smaller than the sum of the halves. This result indicates that the full-length protein is more compact or the linker regions are not as elongated as they are in the separate halves. To further study the possibility that key domain interactions are disrupted when the protein is studied in fragments, we modeled the domains within the *ab initio* envelope by using an atomistic ensemble search method. Overall, the low-grain molecular dynamics describe a tendency for the flexible linkers to expand and contract and for the domains to reorganize themselves in manner that resembles breathing. However, the best-fit structures for the Parp1-1014 or Parp1-486 are somewhat compact and resemble the shape and dimensions of the *ab initio* models of Chapter 3. Interestingly, when the protein is in complex with DNA the molecule is increasingly elongated. Additionally, the models derived from molecular dynamics correspond with the physical parameters that were directly obtained in the various solution studies. The major changes that can be visualized are the extended nature of the linker regions between domains. One linker resides in the region between the BRCT and WGR domains (amino acids 487-528), and contains the most sensitive

trypsin cut-site. We attempted to recapitulate the binding of BRCT and WGR domains but no complexes between Parp1-486 and Parp487-1014 were stable in solution studies (**Supplementary Figure 5.1**). Additionally, the enzymatic function of the enzyme could not be induced even when the critical cofactors (DNA, Mg^{2+} , NAD^+) and the two separate halves of the protein were present. This result suggests that a conformational change within the intact enzyme is crucial for the activation of the catalytic subunit. Examining the electrostatics of the protein do not reveal any obvious regions that might serve as unique charged-patch interfaces upon binding DNA (**Supplementary Figure 5.4**). Perhaps the WGR and BRCT domains do not actually share a stable interface but rather are held in place by one or more specific conformations of the linker domain between them. It is probable that the linker will not perform its intended function, whether it is to relay a signal or serve as a modification target, when not covalently attached to the two flanking domains, as in the full-length protein. In terms of the enzymatic function a recent study by Altmeyer *et al.* using mass spec analysis, elegantly described the major sites of automodification. In that study, the major targets for ADP-ribosylation reside in the very same linker region and specifically include lysine residues K498, K521 and K524 [42]. In our best-fit models of the Parp1-1014 protein alone compared to the DNA bound state, the linker region between the BRCT and the WGR domain expands significantly and the distance between the two domains grows from 40.5Å to 77Å (**Figure 5.20D and E**). Perhaps the non-DNA bound state the sites of automodification remain semi-protected as a form of autoregulation. In such a scenario, the residues would remain slightly buried until the enzyme senses damaged DNA. Upon binding DNA damage, the enzyme would undergo a conformational change and in so doing, the primary target sites for enzymatic attack would then become exposed. Our results describing an increased sensitivity to proteolytic cleavage to residues 494 and 495 in the presence of DNA supports the previous hypothesis. The automodification of critical amino acid targets on the PARP1 protein are a critical step for activating the DNA-damage response of the cell. Regulating the formation the poly(ADP-ribose) polymers might serve as a pathway checkpoint as well. For it is known that the ADP-ribose polymers serve as a signaling mechanism and a binding platform for base excision repair pathway components like XRCC1 and DNA-ligase III [67]. Additionally, from

Chapter 2, the FRET studies of the automodified PARP1 indicate that once activated and modified the protein retains its ability to bind DNA. This newly described ability might serve to anchor the protein to the sites of damage until repair pathway members like Ku70/80 could arrive and protect the ends.

From SAS we found that both Parp1-486 and the full-length PARP-1 undergo large conformational changes when they form complexes with DNA. The cause for the elongation appears to be a binding-induced conformational change that starts at the DBD (amino acids 1-202) but how that event travels through rest of the protein remains unclear however, there is some evidence that the signal can be mediated through the in trans addition of Zn3 . The conformational change described here appears to be due to stretching in the linker regions between domains. Both Parp1-1014 and Parp1-486 undergo an $\sim 20\text{\AA}$ change in R_g and grow in maximum dimension by $\sim 50\text{\AA}$ and $\sim 60\text{\AA}$, respectively. The major support for this claim comes from the SANS contrast variation experiments. The R_g values for the PARP:DNA complexes in the 70% D₂O condition, the DNA match point, are very close to the R_g -values obtained in 0% D₂O and in the SAXS measurements, the conditions where DNA is the dominant scatterer. This Additionally, while building the DNA into the starting models for the complexes, the DNA was moved around to different positions in order to recreate the increase in R_g values. Even when the DNA was moved to the very edge of the N-terminus of our full-length PARP-1 starting structure, an unlikely DNA interface with Zn1, its length could not account for the large differences between the protein and the protein-DNA complexes (**Supplementary Figure 5.3**).

Recently, the mechanism for how the Zn-finger domains bind DNA became more clearly defined in the form of separate crystal structures of Zn1-DNA and Zn2-DNA complexes [36]. Following that publication was one that describes the types of crystallographic interactions between several of the PARP-1 domains and a small oligonucleotide DNA . The crystal structure describes the DNA interactions between the first Zn-finger, Zn1 and the WGR domain, a domain about which little functional knowledge existed. It appears that the Zn3-domain contributes to the interaction by bridging other domain contacts. The authors of the recent structure paper feel that the domain-contacts seen in the crystal structure describe the DNA-induced mechanism of how PARP-1 becomes activated when it binds its allosteric

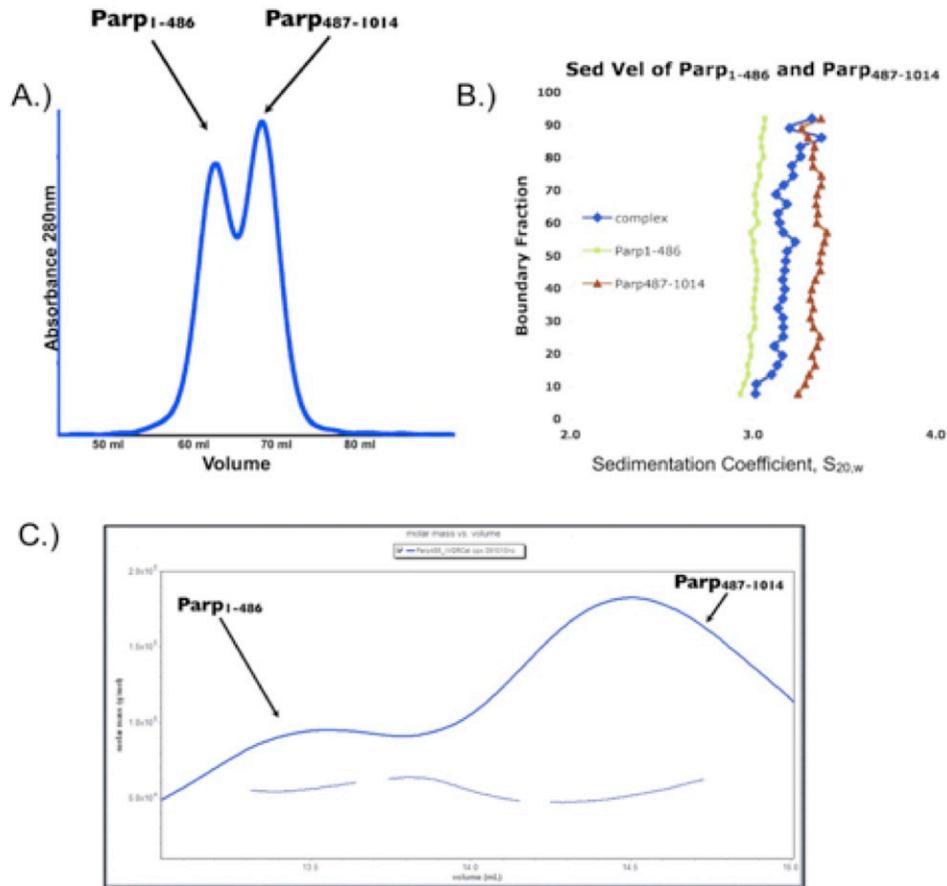
regulator. However, the construct the authors used in that study were missing a large portion of the protein, including the Zn2 domain, the Zn-finger with the greatest affinity for DNA, and the domain that critically mediates contacts between PARP-1 and its XRCC1-complex partners, the BRCT domain. When comparing the calculated scattering profile for the recent PARP-DNA structure PDB file, the R_g -values were 40 Å and 43Å for the protein and the protein-DNA, respectively. Even when the missing domains were added back to the structure the scattering profiles nor the associated R_g -values did not approach our solution scattering R_g -values for the full-length enzyme alone let alone the PARP-1-DNA complex. Regardless of the differences in their results, we have incorporated the various models of DNA-binding in the form of the molecular interactions seen in high-resolution crystal and NMR structures [35, 36, 49] into our modeling as a way to validate and improve our starting structures.

The solution studies described here sheds light on how the entire PARP-1 protein changes when it interacts with DNA. Our studies have focused on how a DNA similar in length to the one used in the crystallization study affects the conformation of the full-length, active human PARP-1 enzyme. This study is the first to describe how the full-length enzyme behaves in solution as a monomer or in complex with a DNA substrate. Our low-resolution studies have described what may be a crucial conformational rearrangement necessary for enzymatic activation.

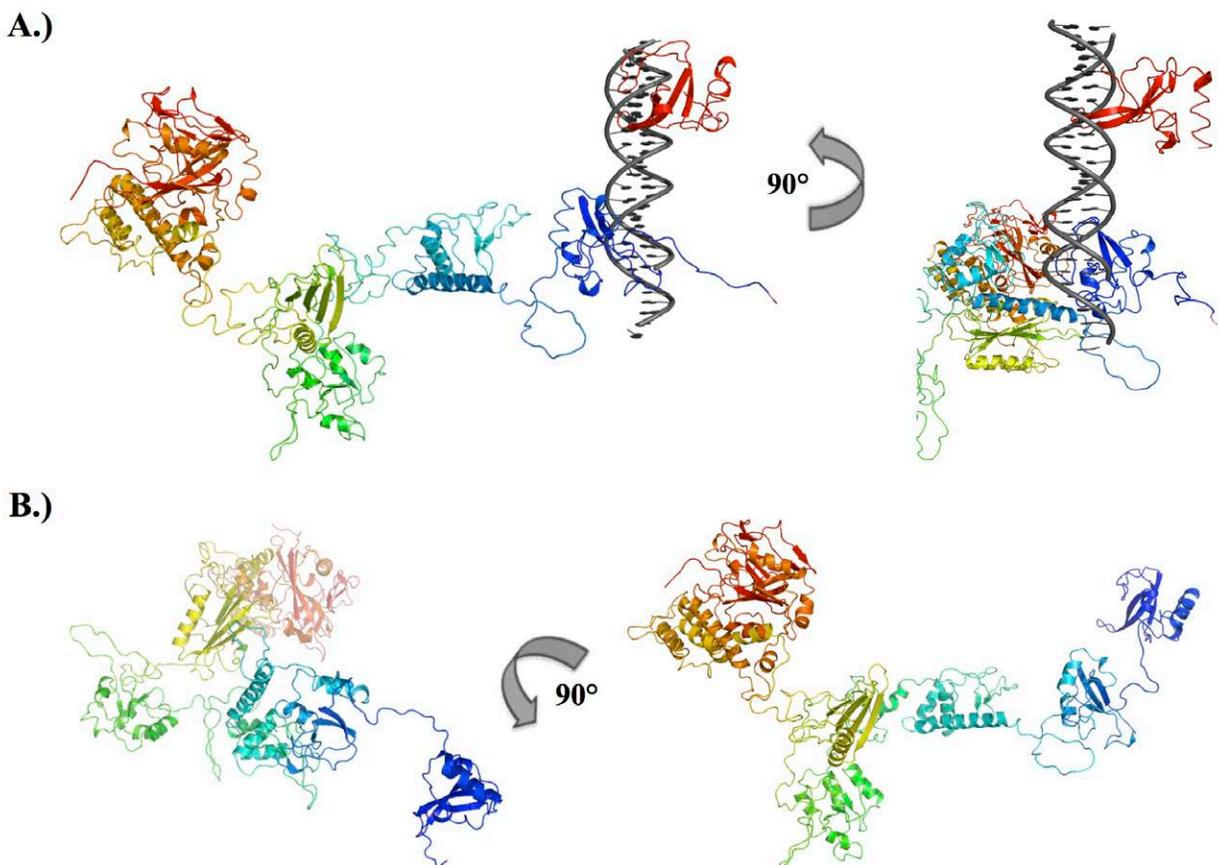
The DNA-induced conformational change of PARP1 described here might serve as a topic of study for future well-designed studies that couple the conformational rearrangement with enzyme kinetics. From such studies it might be possible to better describe the mechanism of activation. Understanding just how the conformational change affects the enzyme at smaller time scales or how those changes influence the actual critical steps of enzymatic functionality will have direct pharmacological value. Case-in-point, different chemotherapeutics can cause different types of DNA damage and many new cancer therapy drugs are in fact PARP-specific inhibitors.

Chapter 5

Supporting Information

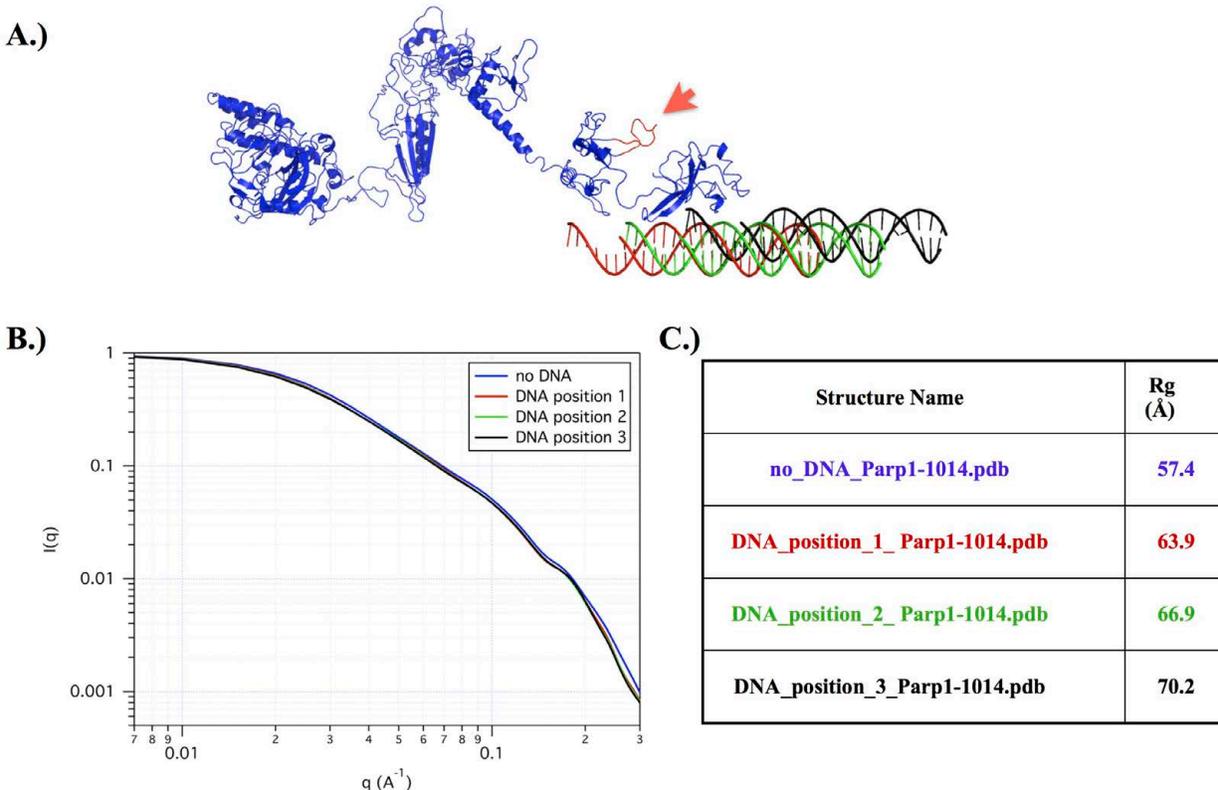


Supplementary Figure 5.1. Solution studies of the interaction of the PARP₁₋₄₈₆ and PARP₄₈₇₋₁₀₁₄ protein fragments. **A.)** Gel filtration experiments on the two halves of PARP-1 incubated at high μM concentrations indicate two components. **B.)** Analytical ultracentrifugation (AUC) sedimentation velocity experiments indicate an average sedimentation coefficient for the “complex” compared to the proteins alone. **C.)** Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) experiments on the proposed complex of PARP₁₋₄₈₆ and PARP₄₈₇₋₁₀₁₄. The molecular mass of the elution profile was monitored continuously and no complex in the expected $\sim 100\text{kDa}$ was observed.

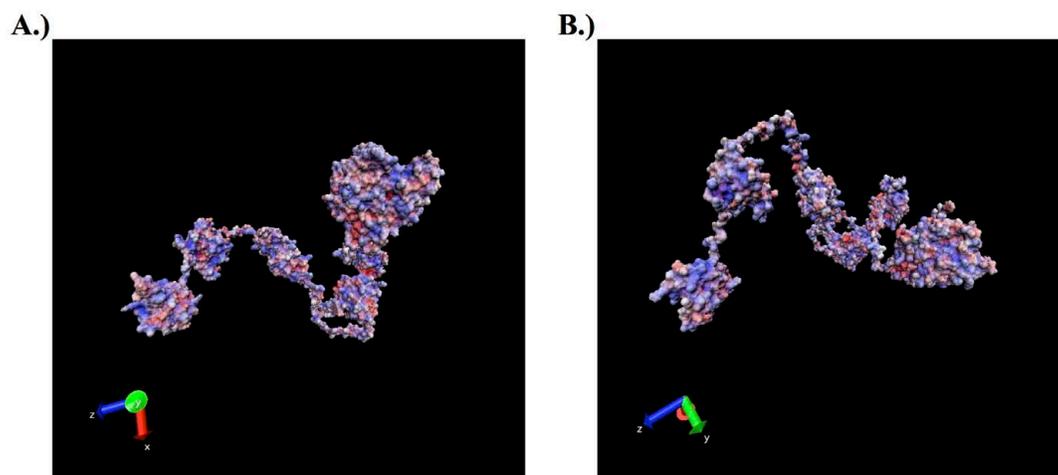


Supplementary Figure 5.2. Atomistic model building for the PARP1:21-27mer 3'-overhang DNA.

A.) The best-fit model for PARP1 was built using the program Modeller (25) as described in the methods section. The PARP1-1014 was then adapted by removing the N-terminal Zn1 domain from the PDB file. The Zn2-DNA structure (pdb 3ODC) was then used to align the Parp98-1014 structure (colored blue to red with a respective N-C terminal gradient) to ensure the proper DNA interface. Next, the PDB of the 21-27mer 3'-overhang was aligned with the DNA used for crystallization. The Zn1 domain (red ribbon) was then inserted into the major groove of the DNA approximately 14bps upstream of the Zn2 binding site (it is known from DNAase protection that the PARP1-1014 has a ~14bp footprint [9]). The binding position is very similar to the position proposed in the crystallographic work by Langelier *et. al.* [18]. Next, the Modeller program was used to re-attach the linker region to the C-terminus of the Zn1-domain. Finally, the entire structure was validated and energy minimized using NAMD and the CHARMM-22 force field. **B.)** The finished, energy minimized structure for the PARP1-DNA complex (the DNA has been removed to better visualize the domain orientations).



Supplementary Figure 5.3. Simulations suggest that placing DNA at the end of the protein cannot reproduce the large change in R_g seen in the experiments of the 21-27bp overhang DNA-PARP1-1014 complex. **A.)** Using the best-fit model of PARP1-1014 ($\chi^2 \sim 1.0$) we tried to recreate the large R_g -change seen in the SANS/SAXS data of the PARP1-1014-DNA complex by placing the DNA as close to the end of the protein as possible. The different models created for this experiment have a different colored DNA molecule. The DNA molecules for each model have a different register, relative to each other. **B.)** Predicted scattering profiles for each of the models with the different DNA-registers (described as frames). At most, the DNA gives a ~ 13 Å increase in R_g (when DNA is bound to the far edge of the protein, a position unlikely to interact with DNA). **C.)** Table of the R_g values derived from the simulated scattering profiles for the models of different DNA register. No matter the register of DNA used in the simulation, the R_g -value does not come close to the real values seen in the experimental data. The crystal structures and binding affinities of Langelier *et. al.*, suggest that Zn2 has the highest affinity for DNA and that the major interface DNA-protein interface occurs at the intercalating loop (colored red and denoted by an orange arrow)[18]. The contrast variation data suggests that even when the DNA signal is negligible, the R_g of the protein in the PARP-DNA complex is ~ 20 Å larger than the protein alone under the same condition. Taken together, the R_g -values from the different contrast conditions point to a large DNA-dependent conformational change.



Supplementary Figure 5.4. The calculated vacuum electrostatic potential for the PARP1-1014 surface residues. The program APBS uses the Poisson-Boltzmann equation to calculate the surface electrostatic potentials for macromolecules [32]. The APBS program can be used in conjunction with the Virtual Molecular Dynamics (VMD) program. The electrostatics were colored red to blue to represent basic and acidic amino acid side chains, respectively. The range of charges plotted was from -30 to 30 kT/e. The protein is predominantly basic in nature and has a predicted pI of ~10. **A.)** and **B.)** are different representative images that are rotated 90° relative to one another. Overall, the charges appear to be homogeneously spread around the surface with no particularly obvious cluster of negative or positive residues with the exception of the Zn1 and Zn2 residues. The basic clusters on the DNA-binding region are complementary to the acidic DNA backbone.

Chapter 6

Summary of Results

Discussion

The experiments herein represent our efforts to characterize PARP-1, thereby contributing to a better basis for how this important protein functions in key biological processes. We focused primarily on determining how PARP-1 functions in the presence of its two major substrates, DNA damage and nucleosomes. The second and fourth chapters relate the relative enzymatic, stoichiometric and thermodynamic parameters of PARP-1 in the presence of these two substrates. We have shown that PARP-1 is a monomer in solution, and it remains so even in complex with its DNA-damage or chromatin substrates. Additionally, PARP-1 as an intact protein has a higher binding affinity for both DNA-damage and chromatin substrates. These findings imply that the N-terminal DNA-binding region may interact with DNA more effectively if its other domains are present, a conclusion consistent with the most recent crystallographic studies of a nearly full-length PARP-1 in complex with a DNA-damage substrate [49].

The work presented in the third, fourth and fifth chapters represent our attempts to provide structural information on the full-length enzyme both alone and in the presence of its DNA-damage substrates. Our experiments are the first to extensively describe the hydrodynamic properties of monodisperse solutions of PARP-1 and PARP-DNA complexes. We have incorporated the independently obtained structural information for the individual domains of PARP-1 and used them to better describe the average solution structure of the full-length enzyme. We have done so by combining atomistic models, molecular dynamics and small-angle scattering techniques to assign experimental constraints and strengthen the otherwise low-resolution structural information. We find that PARP-1 is indeed extremely flexible and can occupy a vast configuration space where the radii of gyration (R_g s) can reside between 45 and 140 Å yet, the best-fit configurations are ~ 60 Å. This finding suggests that, although capable of

extending to lengths where the R_g -values are 140 Å, the protein prefers to reside in a somewhat constrained conformational space. Additionally, from the solution and structural studies we have shown that PARP-1 undergoes a conformational rearrangement in the presence of DNA damage models, a finding that was independently corroborated recently [49].

It is our hope that these studies offer the field a better understanding of how PARP-1 behaves in solution and a basis to design other pertinent experiments. Additionally, the technical and experimental advances we have made will facilitate future efforts to obtain valuable solution information on large, flexible systems that are similar to PARP-1.

Chapter 7

Future Directions

Summary

The body of research presented in this thesis represents my efforts to contribute to the understanding of how PARP-1 performs within several key biochemical systems. As is the case with many studies of complex systems, successful interpretation of the biological problem requires that one first ask the correct questions and then assemble or synthesize the right tools in order to answer those questions. The experiments described here were designed to answer specific questions related to the biological functions of PARP-1. Along the way certain avenues for future experiments have opened up. The following section of this thesis is dedicated to the design of a key experiment I believe will contribute to the understanding of just how PARP-1 binds the nucleosome. The experimental design is not unique, in fact it is very similar to an experiment that helped researchers define a low-resolution description of the nucleosome that was later verified by the crystal structure. The method of choice for study of the nucleosome-PARP-1 interaction would be Small-Angle Neutron Scattering (SANS) in combination with contrast variation. Given the flexibility of PARP-1, and its inherent protease sensitivity, and the need to bind nucleosomal linker DNA to form a proper complex with chromatin, the likelihood that crystallization of the PARP-nucleosome complex will readily occur is quite low. Instead, SANS analysis in conjunction with isotopic substitution (ie contrast variation), of either the histone components or the PARP-1 protein fragments chosen for that study, combined with modern molecular dynamics modeling techniques will help to determine elucidate the binding mode of the PARP-1 to the mononucleosome. The resulting structural information obtained will help to elucidate how PARP-1 binds both nucleosome components as well as the extra-nucleosomal DNA.

Significance:

The enzyme Poly (ADP-Ribose) polymerase-1 (PARP1) is the most abundant member of the PARP-superfamily of proteins[123]. PARP1 performs key roles in two major cellular processes, regulation of chromatin architecture[109, 250] and the sensing of DNA damage[23]. PARP1 binds to nucleosomes and poly (ADP-ribosylates) histones and chromatin-associated proteins to expose specific DNA sequences to the cellular machinery involved in gene transcription and/or DNA damage repair. *In vitro* studies have thus far reported that (a) in the absence of DNA damage, PARP1 is closely associated with silent chromatin and is enzymatically inactive. (b) Upon sensing DNA damage, the enzymatic activity of PARP1 increases dramatically. One key question is how PARP1 is activated upon sensing DNA damage, and how it interacts with chromatin.

We **hypothesized** that the orientation and overall structure of PARP1 when in complex with damaged DNA is different from that of an undamaged chromatin substrate and that this accounts for the activation of the PARP1 enzymatic domain upon interaction with damaged DNA. Here we propose to use SANS to determine the mechanism of binding of PARP1 to the base unit of chromatin, the mononucleosome. It is our hope that these data will lead to an increased understanding of how PARP1 contributes to chromatin architecture.

Preliminary Data:

Human PARP-1 is a modular protein organized into at least 6 domains [211] (**Figure 7.1**). Its affinity for DNA is regulated by two unique zinc-finger motifs (residues 1-209; referred to as zf-PARP, **Figure 7.1**, which are sufficient to target the entire protein to damaged DNA and nucleosomes [124, 251]. Recent publications have focused on a different mode of DNA binding, the PARP1-dependent chromatin binding and subsequent higher-order chromatin compaction [123, 124]. *In vitro* evidence suggests that the zf-PARP fragment can bind to nucleosomes but is inefficient in compacting chromatin when compared to

the full-length protein [124]. It is thought that the C-terminal enzymatic domain of PARP1 with no known DNA-binding ability may contribute to chromatin compaction but it can do so unless it is covalently attached to the N-terminal zf-PARP domain. Since the enzymatic domain is required for compaction of chromatin a specific orientation for PARP1 binding may be required and should be investigated further.

We have already characterized several PARP1 constructs (the full-length included) ([231], unpublished data). Although no multi-domain structures of PARP1 are known, the structure of over 90% of human PARP1 has been defined within six individual domains that have been deposited in the protein databank (PDBID: 2JVN,2DMJ,2CR9,2CS2,2COK,3IID **Figure 7.1B**). Much structural information is available for the nucleosome (NCP) including many high-resolution crystal structures (PDBID: 1AIO **Figure 7.1C**). Our group has begun to study nucleosomes with additional DNA length because it has been shown that PARP1 binds more than the minimum 146 bp needed to form the nucleosome core particle (NCP)[123]. Light scattering and SAXS data shows that full-length PARP1 forms a 1:1 complex with a 206 bp nucleosome (**Figure 7.2A**). Using the known domain structures along with molecular dynamic simulations and ensemble optimized methods we have reconstructed detailed structures of the PARP1 enzyme and 206 bp nucleosome from our data (**Figure 7.1B**). Preliminary studies of how PARP1 binds the “undamaged” model, the nucleosome, are currently underway but require SANS experiments to allow us to distinguish protein from DNA.

In order to maximize our opportunity to collect the high quality SANS data of our protein-DNA complexes, we have established a collaboration with Dr. Susan Krueger, at NIST Center for Neutron research. Using the sequence information for PARP1 and a 206 bp nucleosome, we have designed appropriate SANS D2O contrast protocols (described below) that will allow us to determine protein and DNA components within nucleosomes and PARP1-nucleosome complexes. We anticipate the need for 3 to 5 milligrams of protein per SANS measurement, an amount well within our production capabilities.

To date, we have begun testing the solubility of our PARP1 constructs in increasing amounts of D2O and are optimizing buffer conditions in which our PARP1 samples are sufficiently stable for SANS

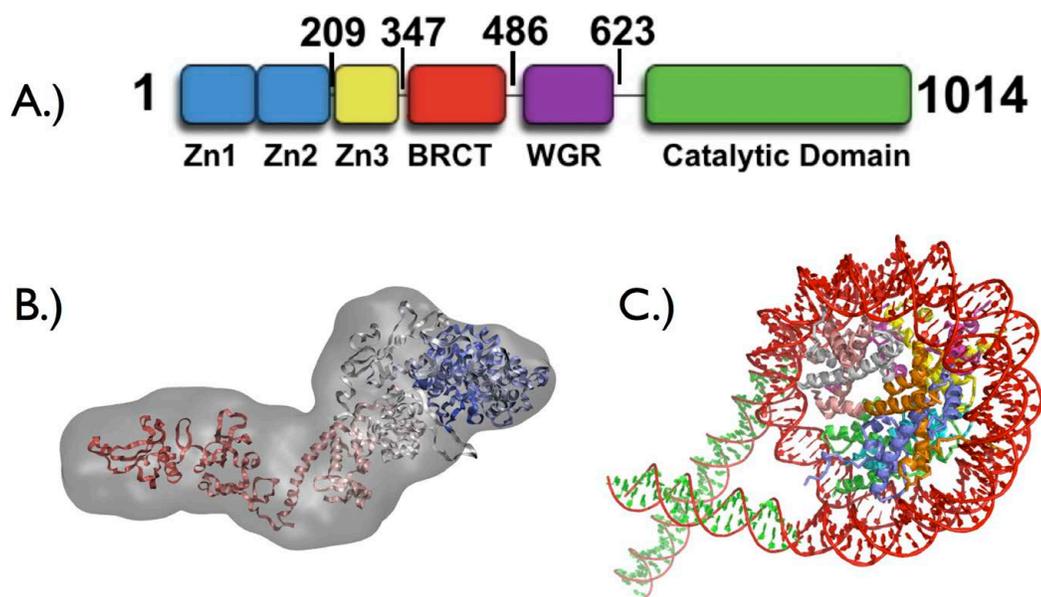


Figure 7.1: Starting models for the study of human PARP1-nucleosome complex formation.

A.) The human PARP1 protein contains six known domains. The two N-terminal zinc fingers (Zn1, Zn2) are responsible for binding damaged DNA in a process that activates the C-terminal catalytic domain. **B.)** A Model of the tertiary structure of full-length Human PARP1. The *Ab initio* calculated envelop of full-length PARP1 is illustrated in grey, while the tertiary structure of the protein is represented in a red-blue gradient (N-C terminus) The tertiary structure of PARP1 was generated by rigid-body fitting of individual domains using the SASREF program building linker regions between domains and refining the structure by energy minimization and position restrained refinement, (Initial atomistic PARP-1 model was generated by Wayne Lilyestrom PhD and it is not to scale with the nucleosome) **C.)** A model of a 206 bp nucleosome based on the nucleosome crystal structure, 1A10. Green bases contained within red phosphate backbones designate the linker DNA not required for formation of the nucleosome core particle (NCP).

analysis. Finally we have recently received rapid access SAXS time at SSRL beamline BL4-2 in and we have collected data of protein-DNA complexes in both H₂O and D₂O. Light scattering and SAXS data suggest differences in scattering profiles are likely due to conformational changes of the PARP1-nucleosome complex and not aggregation (**Figure 7.2B**). In conclusion, SANS analysis of PARP1-nucleosome complexes will provide key information on how PARP1 functions in the presence of chromatin and possibly lead to a better understanding of how PARP1 functions when bound to its other substrate, damaged DNA.

Experimental Plan:

We would ask for three days of beamtime to perform a complete contrast variation series of experiments on the nucleosome and the PARP1-nucleosome complex. The 206 bp nucleosome will allow us to determine the trajectory of the linker DNA (currently unknown). The PARP1-nucleosome complex will allow the determination of the structure of the PARP1 protein in the complex (for comparison to our previously obtained SAXS data). We plan to measure the complex under buffer conditions with 0%, 15%, ~40%, ~65%, 85% and 100% D₂O. Thus, 3 measurements will be made under each contrast condition: 206 bp nucleosome, PARP1-nucleosome complex and buffer background. Thus, 3 measurements will be made under each contrast condition: 206 bp nucleosome, PARP1-nucleosome complex and buffer background. The same buffer background measurement will be used for both complexes at a given contrast condition. The scattering from the PARP1-nucleosome sample may be strong enough to measure in 30-40% D₂O buffer to obtain data as close to the protein match point as possible. At least two instrument configurations will be needed in order to cover a q-range large enough to encompass the range of R_g values expected as a function of contrast. While the R_g values of the nucleosome alone are expected to be below ~60Å, significantly larger R_g values are expected for the PARP1-nucleosome

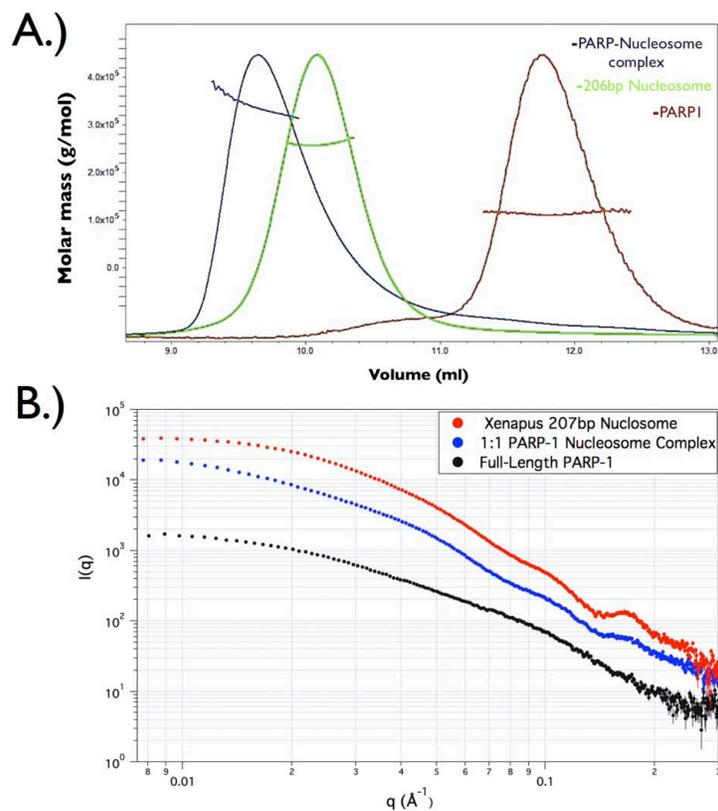


Figure 7.2. Biophysical studies of the PARP1-nucleosome complex. A.) Size exclusion chromatography coupled multi-angle light scattering (SEC-MALS) experiments of PARP1, 206 bp nucleosome and the 1:1 PARP-nucleosome complex (110KDa, 230KDa and 340KDa, respectively). Titration experiments show PARP1 prefers to form a 1:1 complex with nucleosomes even though two free DNA ends, known PARP1 recognition sites, are present (see **Chapter 2** for more details). B.) Small angle light scattering (SAXS) profiles of the 206 bp nucleosome and PARP-nucleosome complex.

complex, ($>85\text{\AA}$). The protein concentration will be approximately 4 mg/ml in the two samples. A lower concentration will be used for the higher-D₂O contrast conditions, if necessary. In Dr. Krueger's experience, it will take three days of beamtime to perform these two simultaneous contrast variation series of measurements.

SANS data will be analyzed initially using the standard Guinier analysis to determine R_g and $I(0)$. These parameters will also be determined using standard $P(r)$ analysis. We will plot $\sqrt{I(0)}$ vs %D₂O in order to verify the match point of the complexes. We will also perform a Stuhrmann analysis to determine R_g for the protein and DNA components in the complex as well as the separation of the centers of masses of the two components. Software to perform these analyses exists at the NCNR. We will also use the web-based MULCH software[252] to perform these analyses and to calculate the separate intensities for the protein and DNA in the complexes. Atomistic structure models of the separate components and the entire complex will be determined using the SASSIE suite of software developed at the NCNR for the modeling of flexible proteins in solution, using the starting structures described in Figure 6.1. SASSIE uses a combination of Monte Carlo and molecular dynamics approaches to find a family of structures consistent with the SANS data.

References:

1. Franklin, R.E. and R.G. Gosling, *Molecular configuration in sodium thymonucleate*. Nature, 1953. **171**(4356): p. 740-741.
2. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
3. Kendrew, J.C., et al., *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. Nature, 1958. **181**(4610): p. 662-6.
4. Chambon, P., J.D. Weill, and P. Mandel, *Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme*. Biochem Biophys Res Commun, 1963. **11**: p. 39-43.
5. Hilz, H. and P. Stone, *Poly(ADP-ribose) and ADP-ribosylation of proteins*. Rev Physiol Biochem Pharmacol, 1976. **76**: p. 1-58, 177.
6. Chambon, P., M. Ramuz, and J. Doly, *Relation between soluble DNA-dependent RNA polymerase and "aggregate" RNA polymerase*. Biochem Biophys Res Commun, 1965. **21**(2): p. 156-61.
7. Nishizuka, Y., et al., *Studies on the polymer of adenosine diphosphate ribose. I. Enzymic formation from nicotinamide adenine dinucleotide in mammalian nuclei*. J Biol Chem, 1967. **242**(13): p. 3164-71.
8. Reeder, R.H., et al., *Studies on the polymer of adenosine diphosphate ribose. II. Characterization of the polymer*. J Biol Chem, 1967. **242**(13): p. 3172-9.
9. Ueda, K., et al., *Poly ADP-ribose glycohydrolase from rat liver nuclei, a novel enzyme degrading the polymer*. Biochem Biophys Res Commun, 1972. **46**(2): p. 516-23.
10. Nishizuka, Y., et al., *Enzymic adenosine diphosphate ribosylation of histone and poly adenosine diphosphate ribose synthesis in rat liver nuclei*. J Biol Chem, 1968. **243**(13): p. 3765-7.
11. Messner, S., et al., *PARP1 ADP-ribosylates lysine residues of the core histone tails*. Nucleic Acids Res, 2010. **38**(19): p. 6350-62.
12. Panzeter, P.L., C.A. Realini, and F.R. Althaus, *Noncovalent interactions of poly(adenosine diphosphate ribose) with histones*. Biochemistry, 1992. **31**(5): p. 1379-85.
13. Althaus, F.R., et al., *Histone shuttle driven by the automodification cycle of poly(ADP-ribose)polymerase*. Environ Mol Mutagen, 1993. **22**(4): p. 278-82.
14. Lepiniec, L., et al., *Characterization of an Arabidopsis thaliana cDNA homologue to animal poly(ADP-ribose) polymerase*. FEBS Lett, 1995. **364**(2): p. 103-8.
15. Babiychuk, E., et al., *Higher plants possess two structurally different poly(ADP-ribose) polymerases*. Plant J, 1998. **15**(5): p. 635-45.
16. Shieh, W.M., et al., *Poly(ADP-ribose) polymerase null mouse cells synthesize ADP-ribose polymers*. J Biol Chem, 1998. **273**(46): p. 30069-72.
17. Ame, J.C., et al., *PARP-2, A novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase*. J Biol Chem, 1999. **274**(25): p. 17860-8.
18. Menissier de Murcia, J., et al., *Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse*. Embo J, 2003. **22**(9): p. 2255-63.
19. Smith, S., et al., *Tankyrase, a poly(ADP-ribose) polymerase at human telomeres*. Science, 1998. **282**(5393): p. 1484-7.

20. Johansson, M., *A human poly(ADP-ribose) polymerase gene family (ADPRTL): cDNA cloning of two novel poly(ADP-ribose) polymerase homologues*. Genomics, 1999. **57**(3): p. 442-5.
21. Augustin, A., et al., *PARP-3 localizes preferentially to the daughter centriole and interferes with the G1/S cell cycle progression*. J Cell Sci, 2003. **116**(Pt 8): p. 1551-62.
22. Boehler, C., et al., *Poly(ADP-ribose) polymerase 3 (PARP3), a newcomer in cellular response to DNA damage and mitotic progression*. Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2783-8.
23. Ame, J.C., C. Spenlehauer, and G. de Murcia, *The PARP superfamily*. Bioessays, 2004. **26**(8): p. 882-93.
24. Diefenbach, J. and A. Burkle, *Introduction to poly(ADP-ribose) metabolism*. Cell Mol Life Sci, 2005. **62**(7-8): p. 721-30.
25. Kleine, H., et al., *Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation*. Mol Cell, 2008. **32**(1): p. 57-69.
26. Hottiger, M.O., et al., *Toward a unified nomenclature for mammalian ADP-ribosyltransferases*. Trends Biochem Sci, 2010. **35**(4): p. 208-19.
27. Ikai, K., K. Ueda, and O. Hayaishi, *Immunohistochemical demonstration of poly(adenosine diphosphate-ribose) in nuclei of various rat tissues*. J Histochem Cytochem, 1980. **28**(7): p. 670-6.
28. Yoshihara, K., et al., *Bovine thymus poly(adenosine diphosphate ribose) polymerase*. J Biol Chem, 1978. **253**(18): p. 6459-66.
29. Ueda, K. and O. Hayaishi, *ADP-ribosylation*. Annu Rev Biochem, 1985. **54**: p. 73-100.
30. Lamarre, D., et al., *Production and characterization of monoclonal antibodies specific for the functional domains of poly(ADP-ribose) polymerase*. Biochem Cell Biol, 1986. **64**(4): p. 368-76.
31. Stanker, L.H., M. Vanderlaan, and H. Juarez-Salinas, *One-step purification of mouse monoclonal antibodies from ascites fluid by hydroxylapatite chromatography*. J Immunol Methods, 1985. **76**(1): p. 157-69.
32. Zahradka, P. and K. Ebisuzaki, *Poly(ADP-ribose) polymerase is a zinc metalloenzyme*. Eur J Biochem, 1984. **142**(3): p. 503-9.
33. Menissier-de Murcia, J., et al., *Zinc-binding domain of poly(ADP-ribose) polymerase participates in the recognition of single strand breaks on DNA*. J Mol Biol, 1989. **210**(1): p. 229-33.
34. Initiative, R.S.G.a.P., . ongoing
35. Eustermann, S., et al., *The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger*. J Mol Biol, 2011. **407**(1): p. 149-70.
36. Langelier, M.F., et al., *Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity*. J Biol Chem, 2011(1083-351X (Electronic)).
37. Kaufmann, S.H., et al., *Specific proteolytic cleavage of poly(ADP-ribose) polymerase: an early marker of chemotherapy-induced apoptosis*. Cancer Res, 1993. **53**(17): p. 3976-85.
38. Lazebnik, Y.A., et al., *Cleavage of poly(ADP-ribose) polymerase by a proteinase with properties like ICE*. Nature, 1994. **371**(6495): p. 346-7.
39. Beernink, P.T., et al., *Specificity of protein interactions mediated by BRCT domains of the XRCC1 DNA repair protein*. J Biol Chem, 2005. **280**(34): p. 30206.

40. Loeffler, P.A., et al., *Structural studies of the PARP-1 BRCT domain*. BMC Struct Biol, 2011. **11**: p. 37.
41. Hassa, P.O., et al., *Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going?* Microbiol Mol Biol Rev, 2006. **70**(3): p. 789-829.
42. Altmeyer, M., et al., *Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites*. Nucleic Acids Res, 2009.
43. Ruf, A., et al., *Structure of the catalytic fragment of poly(AD-ribose) polymerase from chicken*. Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7481-5.
44. Ruf, A., et al., *The mechanism of the elongation and branching reaction of poly(ADP-ribose) polymerase as derived from crystal structures and mutagenesis*. J Mol Biol, 1998. **278**(1): p. 57-65.
45. Ohno, K., et al., *Docking study and binding free energy calculation of poly (ADP-ribose) polymerase inhibitors*. J Mol Model, 2011. **17**(2): p. 383-9.
46. Cipak, L. and S. Jantova, *PARP-1 inhibitors: a novel genetically specific agents for cancer therapy*. Neoplasma, 2010. **57**(5): p. 401-5.
47. Tao, Z., et al., *Domain C of human poly(ADP-ribose) polymerase-1 is important for enzyme activity and contains a novel zinc-ribbon motif*. Biochemistry, 2008. **47**(21): p. 5804-13.
48. Langelier, M.F., et al., *A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation*. J Biol Chem, 2008. **283**(7): p. 4105-14.
49. Langelier, M.F., et al., *Structural basis for DNA damage-dependent poly(ADP-ribosylation) by human PARP-1*. Science, 2012. **336**(6082): p. 728-32.
50. Dyson, H.J., *Expanding the proteome: disordered and alternatively folded proteins*. Q Rev Biophys, 2011. **44**(4): p. 467-518.
51. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
52. Tainer, J.A., et al., *The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein*. Nature, 1984. **312**(5990): p. 127-34.
53. Sandhu, K.S., *Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins*. J Mol Recognit, 2009. **22**(1): p. 1-8.
54. Ohgushi, H., K. Yoshihara, and T. Kamiya, *Bovine thymus poly(adenosine diphosphate ribose) polymerase. Physical properties and binding to DNA*. J Biol Chem, 1980. **255**(13): p. 6205-11.
55. Kinoshita, T., et al., *Inhibitor-induced structural change of the active site of human poly(ADP-ribose) polymerase*. FEBS Lett, 2004. **556**(1-3): p. 43-6.
56. Sallmann, F.R., et al., *Characterization of sPARP-1. An alternative product of PARP-1 gene with poly(ADP-ribose) polymerase activity independent of DNA strand breaks*. J Biol Chem, 2000. **275**(20): p. 15504-11.
57. Simonin, F., et al., *Identification of potential active-site residues in the human poly(ADP-ribose) polymerase*. J Biol Chem, 1993. **268**(12): p. 8529-35.
58. Berger, N.A., et al., *Poly(ADP-ribose) polymerase mediates the suicide response to massive DNA damage: studies in normal and DNA-repair defective cells*. Princess Takamatsu Symp, 1983. **13**: p. 219-26.

59. Gradwohl, G., et al., *The second zinc-finger domain of poly(ADP-ribose) polymerase determines specificity for single-stranded breaks in DNA*. Proc Natl Acad Sci U S A, 1990. **87**(8): p. 2990-4.
60. de Murcia, G. and J. Menissier de Murcia, *Poly(ADP-ribose) polymerase: a molecular nick-sensor*. Trends Biochem Sci, 1994. **19**(4): p. 172-6.
61. Le Cam, E., et al., *Conformational analysis of a 139 base-pair DNA fragment containing a single-stranded break and its interaction with human poly(ADP-ribose) polymerase*. J Mol Biol, 1994. **235**(3): p. 1062-1071.
62. Eustermann, S., et al., *The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger*. J. Mol Biol., 2011. **407**(1): p. 149-170.
63. Leppard, J.B., et al., *Physical and functional interaction between DNA ligase IIIalpha and poly(ADP-Ribose) polymerase I in DNA single-strand break repair*. Mol Cell Biol, 2003. **23**(16): p. 5919-27.
64. Le Page, F., et al., *Poly(ADP-ribose) polymerase-1 (PARP-1) is required in murine cell lines for base excision repair of oxidative DNA damage in the absence of DNA polymerase beta*. J Biol Chem, 2003. **278**(20): p. 18471-7.
65. Knight, M.I. and P.J. Chambers, *Production, extraction, and purification of human poly(ADP-ribose) polymerase-1 (PARP-1) with high specific activity*. Protein Expr Purif, 2001. **23**(3): p. 453-8.
66. Audebert, M., B. Salles, and P. Calsou, *Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining*. J Biol Chem, 2004. **279**(53): p. 55117-26.
67. Pleschke, J.M., et al., *Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins*. J Biol Chem, 2000. **275**(52): p. 40974-80.
68. Masson, M., et al., *XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage*. Mol Cell Biol, 1998. **18**(6): p. 3563-71.
69. Tong, W.M., U. Cortes, and Z.Q. Wang, *Poly(ADP-ribose) polymerase: a guardian angel protecting the genome and suppressing tumorigenesis*. Biochim Biophys Acta, 2001. **1552**(1): p. 27-37.
70. D'Silva, I., et al., *Relative affinities of poly(ADP-ribose) polymerase and DNA-dependent protein kinase for DNA strand interruptions*. Biochim Biophys Acta, 1999. **1430**: p. 119-126.
71. Morrison, C., et al., *Genetic interaction between PARP and DNA-PK in V(D)J recombination and tumorigenesis*. Nat Genet, 1997. **17**(4): p. 479-82.
72. Galande, S. and T. Kohwi-Shigematsu, *Poly(ADP-ribose) polymerase and Ku autoantigen form a complex and synergistically bind to matrix attachment sequences*. J. Biol. Chem., 1999. **274**(29): p. 20521-20528.
73. Brown, M.L., et al., *Role of poly(ADP-ribosyl)ation in DNA-PKcs- independent V(D)J recombination*. Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4532-7.
74. Paddock, M.N., et al., *Competition between PARP-1 and Ku70 control the decision between high-fidelity and mutagenic DNA repair*. DNA Repair (Amst), 2011. **10**(3): p. 338-43.
75. Spagnolo, L., et al., *Visualization of a DNA-PK/PARP1 complex*. Nucleic Acids Res, 2012. **40**(9): p. 4168-77.

76. Hammel, M., et al., *Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex*. J Biol Chem, 2010. **285**(2): p. 1414-23.
77. d'Adda di Fagagna, F., et al., *Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability*. Nat Genet, 1999. **23**(1): p. 76-80.
78. Tong, W.M., et al., *DNA strand break-sensing molecule poly(ADP-Ribose) polymerase cooperates with p53 in telomere function, chromosome stability, and tumor suppression*. Mol Cell Biol, 2001. **21**(12): p. 4046-54.
79. Simbulan-Rosenthal, C.M., et al., *Chromosomal aberrations in PARP(-/-) mice: genome stabilization in immortalized cells by reintroduction of poly(ADP-ribose) polymerase cDNA*. Proc Natl Acad Sci U S A, 1999. **96**(23): p. 13191-6.
80. Rouleau, M., et al., *PARP inhibition: PARP1 and beyond*. Nat Rev Cancer, 2010. **10**(4): p. 293-301.
81. Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy*. Nature, 2005. **434**(7035): p. 917-21.
82. Bryant, H.E., et al., *Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase*. Nature, 2005. **434**(7035): p. 913-7.
83. McCabe, N., et al., *BRCA2-deficient CAPAN-1 cells are extremely sensitive to the inhibition of Poly (ADP-Ribose) polymerase: an issue of potency*. Cancer Biol Ther, 2005. **4**(9): p. 934-6.
84. Beneke, S. and A. Burkle, *Poly(ADP-ribosyl)ation in mammalian ageing*. Nucleic Acids Res, 2007.
85. Burkle, A., *PARP-1: a regulator of genomic stability linked with mammalian longevity*. ChemBiochem, 2001. **2**(10): p. 725-8.
86. Burkle, A., *Poly(ADP-ribosyl)ation: a posttranslational protein modification linked with genome protection and mammalian longevity*. Biogerontology, 2000. **1**(1): p. 41-6.
87. Beneke, S., R. Alvarez-Gonzalez, and A. Burkle, *Comparative characterisation of poly(ADP-ribose) polymerase-1 from two mammalian species with different life span*. Exp Gerontol, 2000. **35**(8): p. 989-1002.
88. Muiras, M.L., et al., *Increased poly(ADP-ribose) polymerase activity in lymphoblastoid cell lines from centenarians*. J Mol Med, 1998. **76**(5): p. 346-54.
89. Burkle, A., et al., *Poly(ADP-ribose) polymerase activity in intact or permeabilized leukocytes from mammalian species of different longevity*. Mol Cell Biochem, 1994. **138**(1-2): p. 85-90.
90. Burkle, A., K. Grube, and J.H. Kupper, *Poly(ADP-ribosyl)ation: its role in inducible DNA amplification, and its correlation with the longevity of mammalian species*. Exp Clin Immunogenet, 1992. **9**(4): p. 230-40.
91. Grube, K. and A. Burkle, *Poly(ADP-ribose) polymerase activity in mononuclear leukocytes of 13 mammalian species correlates with species-specific life span*. Proc Natl Acad Sci U S A, 1992. **89**(24): p. 11759-63.
92. Beneke, S., R. Alvarez-Gonzalez, and A. Bürkle, *Comparative characterisation of poly(ADP-ribose) polymerase-1 from two mammalian species with different life span*. Exp Gerontol, 2000. **35**(8): p. 989-1002.
93. Beneke, S., et al., *Enzyme characteristics of recombinant poly(ADP-ribose) polymerases-1 of rat and human origin mirror the correlation between cellular poly(ADP-*

- ribosyl)ation capacity and species-specific life span. Mech Ageing Dev, 2010. 131(5): p. 366-9.*
94. Wang, X.G., et al., *PARP1 Val762Ala polymorphism reduces enzymatic activity. Biochem Biophys Res Commun, 2007. 354(1): p. 122-6.*
 95. Li, C., et al., *Genetic polymorphisms in DNA base-excision repair genes ADPRT, XRCC1, and APE1 and the risk of squamous cell carcinoma of the head and neck. Cancer, 2007. 110(4): p. 867-75.*
 96. Lockett, K.L., et al., *The ADPRT V762A genetic variant contributes to prostate cancer susceptibility and deficient enzyme function. Cancer Res, 2004. 64(17): p. 6344-8.*
 97. Tao, Z., P. Gao, and H.W. Liu, *Identification of the ADP-ribosylation sites in the PARP-1 automodification domain: analysis and implications. J Am Chem Soc, 2009. 131(40): p. 14258-60.*
 98. Kennedy, R.D., et al., *The role of BRCA1 in the cellular response to chemotherapy. J Natl Cancer Inst, 2004. 96(22): p. 1659-68.*
 99. Zhu, G., P. Chang, and S.J. Lippard, *Recognition of platinum-DNA damage by poly(ADP-ribose) polymerase-1. Biochemistry, 2010. 49(29): p. 6177-83.*
 100. Yamanaka, H., et al., *Characterization of human poly(ADP-ribose) polymerase with autoantibodies. J Biol Chem, 1988. 263(8): p. 3879-83.*
 101. Ludwig, A., et al., *Immunoquantitation and size determination of intrinsic poly(ADP-ribose) polymerase from acid precipitates. An analysis of the in vivo status in mammalian species and in lower eukaryotes. J Biol Chem, 1988. 263(15): p. 6993-9.*
 102. Matsui, T., et al., *Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J Biol Chem, 1980. 255(24): p. 11992-6.*
 103. D'Amours, D., et al., *Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. Biochem J, 1999. 342 (Pt 2): p. 249-68.*
 104. Hough, C.J. and M.E. Smulson, *Association of poly(adenosine diphosphate ribosylated) nucleosomes with transcriptionally active and inactive regions of chromatin. Biochemistry, 1984. 23(21): p. 5016-23.*
 105. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature, 1997. 389(6648): p. 251-60.*
 106. Loetscher, P., R. Alvarez-Gonzalez, and F.R. Althaus, *Poly(ADP-ribose) May Signal Changing Metabolic Conditions to the Chromatin of Mammalian Cells. PNAS, 1987. 84(5): p. 1286-1289.*
 107. Kawaichi, M., K. Ueda, and O. Hayaishi, *Multiple autopoly(ADP-ribosyl)ation of rat liver poly(ADP-ribose) synthetase. Mode of modification and properties of automodified synthetase. J Biol Chem, 1981. 256(18): p. 9483-9.*
 108. Bauer, P.I., A. Hakam, and E. Kun, *Mechanisms of poly(ADP-ribose) polymerase catalysis; mono-ADP-ribosylation of poly(ADP-ribose) polymerase at nanomolar concentrations of NAD. FEBS Lett, 1986. 195(1-2): p. 331-8.*
 109. Tulin, A. and A. Spradling, *Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci. Science, 2003. 299(5606): p. 560-2.*
 110. Pinnola, A., et al., *Nucleosomal core histones mediate dynamic regulation of poly(ADP-ribose) polymerase I protein binding to chromatin and induction of its enzymatic activity. J Biol Chem, 2007. 282(44): p. 32511-9.*
 111. Pirrotta, V., *Transcription. Puffing with PARP. Science, 2003. 299(5606): p. 528-9.*
 112. Kraus, W.L. and J.T. Lis, *PARP goes transcription. Cell, 2003. 113(6): p. 677-83.*

113. Cartwright, I.L. and S.C. Elgin, *Nucleosomal instability and induction of new upstream protein-DNA associations accompany activation of four small heat shock protein genes in Drosophila melanogaster*. Mol Cell Biol, 1986. **6**(3): p. 779-91.
114. Cohen-Armon, M., et al., *DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation*. Mol Cell, 2007. **25**(2): p. 297-308.
115. Kauppinen, T.M., et al., *Direct phosphorylation and regulation of poly(ADP-ribose) polymerase-1 by extracellular signal-regulated kinases 1/2*. Proc Natl Acad Sci U S A, 2006. **103**(18): p. 7136-41.
116. Chen, Z., et al., *MAP kinases*. Chem Rev, 2001. **101**(8): p. 2449-76.
117. Aoufouchi, S. and S. Shall, *Regulation by phosphorylation of Xenopus laevis poly(ADP-ribose) polymerase enzyme activity during oocyte maturation*. Biochem J, 1997. **325**(Pt 2): p. 543-51.
118. Bauer, P.I., et al., *Inhibition of DNA binding by the phosphorylation of poly ADP-ribose polymerase protein catalysed by protein kinase C*. Biochem Biophys Res Commun, 1992. **187**(2): p. 730-6.
119. Kotova, E., et al., *Drosophila histone H2A variant (H2Av) controls poly(ADP-ribose) polymerase 1 (PARP1) activation in chromatin*. Proc Natl Acad Sci U S A, 2011. **108**(15): p. 6205-10.
120. Gradwohl, G., A. Mazen, and G. de Murcia, *Poly(ADP-ribose) polymerase forms loops with DNA*. Biochem Biophys Res Commun, 1987. **148**(No. 3): p. 913-919.
121. De Bernardin, W., R. Losa, and T. Koller, *Formation and characterization of soluble complexes of histone H1 with supercoiled DNA*. J Mol Biol, 1986. **189**(3): p. 503-17.
122. Ivanchenko, M., J. Zlatanova, and K. van Holde, *Histone H1 preferentially binds to superhelical DNA molecules of higher compaction*. Biophys J, 1997. **72**(3): p. 1388-95.
123. Kim, M.Y., et al., *NAD⁺-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1*. Cell, 2004. **119**(6): p. 803-14.
124. Wacker, D.A., et al., *The DNA Binding and Catalytic Domains of Poly(ADP-ribose) Polymerase-1 Cooperate in the Regulation of Chromatin Structure and Transcription*. Mol Cell Biol, 2007.
125. Langelier, M.F., et al., *The Zn³ domain of human poly(ADP-ribose) polymerase-1 (PARP-1) functions in both DNA-dependent poly(ADP-ribose) synthesis activity and chromatin compaction*. J Biol Chem, 2010. **285**(24): p. 18877-87.
126. Anders, C.K., et al., *Poly(ADP-Ribose) polymerase inhibition: "targeted" therapy for triple-negative breast cancer*. Clin Cancer Res., 2010. **16**(19): p. 4702-10.
127. Kraus, W.L., *Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation*. Curr Opin Cell Biol, 2008. **20**(3): p. 294-302.
128. Lord, C.J. and A. Ashworth, *The DNA damage response and cancer therapy*. Nature, 2012. **481**(7381): p. 287-94.
129. Krishnakumar, R. and W.L. Kraus, *The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets*. Mol Cell, 2010. **39**(1): p. 8-24.
130. Ji, Y. and A.V. Tulin, *The roles of PARP1 in gene control and cell differentiation*. Curr Opin Genet Dev, 2010. **20**(5): p. 512-8.
131. Krishnakumar, R., et al., *Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes*. Science, 2008. **319**(5864): p. 819-21.
132. Gibson, B.A. and W.L. Kraus, *New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs*. Nat Rev Mol Cell Biol, 2012.

133. Altmeyer, M., et al., *Molecular mechanism of poly(ADP-ribosylation) by PARP1 and identification of lysine residues as ADP-ribose acceptor sites*. Nucleic Acids Res, 2009. **37**(11): p. 3723-38.
134. Mangerich, A. and A. Burkle, *How to kill tumor cells with inhibitors of poly(ADP-ribosylation)*. Int J Cancer, 2011. **128**(2): p. 251-65.
135. Javle, M. and N.J. Curtin, *The role of PARP in DNA repair and its therapeutic exploitation*. Br J Cancer, 2011. **105**(8): p. 1114-22.
136. Lilyestrom, W., et al., *Structural and biophysical studies of human PARP-1 in complex with damaged DNA*. J Mol Biol, 2010. **395**(5): p. 983-94.
137. Langelier, M.F., et al., *Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity*. J Biol Chem, 2011.
138. Winkler, D.D. and K. Luger, *The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization*. J Biol Chem, 2011. **286**(21): p. 18369-74.
139. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning*. J Mol Biol, 1998. **276**(1): p. 19-42.
140. Schreiber, V., et al., *The human poly(ADP-ribose) polymerase nuclear localization signal is a bipartite element functionally separate from DNA binding and catalytic activity*. Embo J, 1992. **11**(9): p. 3263-9.
141. Wacker, D.A., et al., *The DNA binding and catalytic domains of poly(ADP-ribose) polymerase 1 cooperate in the regulation of chromatin structure and transcription*. Mol Cell Biol, 2007. **27**(21): p. 7475-85.
142. Hieb, A.R., et al., *Fluorescence strategies for high-throughput quantification of protein interactions*. Nucleic Acids Res, 2012. **40**(5): p. e33.
143. Dyer, P.N., et al., *Reconstitution of nucleosome core particles from recombinant histones and DNA*. Methods Enzymol, 2004. **375**: p. 23-44.
144. Yang, C., et al., *Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes*. Nucleic Acids Res, 2011. **39**(10): p. 4122-35.
145. Winkler, D.D., K. Luger, and A.R. Hieb, *Quantifying chromatin associated interactions: the HI-FI System*. Methods Enzymol, 2012. **in press**.
146. Koo, H.S., H.M. Wu, and D.M. Crothers, *DNA bending at adenine . thymine tracts*. Nature, 1986. **320**(6062): p. 501-6.
147. Li, G. and J. Widom, *Nucleosomes facilitate their own invasion*. Nat Struct Mol Biol., 2004. **11**: p. 763-769.
148. Gurunathan, K. and M. Levitus, *Single-molecule fluorescence studies of nucleosome dynamics*. Curr Pharm Biotechnol, 2009. **10**(5): p. 559-68.
149. Yang, W., *Poor base stacking at DNA lesions may initiate recognition by many repair proteins*. DNA Repair (Amst), 2006. **5**(6): p. 654-66.
150. Gagne, J.P., M. Rouleau, and G.G. Poirier, *Structural biology. PARP-1 activation--bringing the pieces together*. Science, 2012. **336**(6082): p. 678-9.
151. Ali, A.A., et al., *The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks*. Nat Struct Mol Biol, 2012.
152. Kim, M.Y., T. Zhang, and W.L. Kraus, *Poly(ADP-ribosylation) by PARP-1: 'PAR-laying' NAD+ into a nuclear signal*. Genes Dev, 2005. **19**(17): p. 1951-67.

153. David, M.B., O.N. Brik, and A. Paz, *Assessment of CASP8 structure predictions for template free targets*. Proteins: Struct, Funct, Bioinf, 2009. **77**: p. 50-65.
154. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution*. Q Rev Biophys, 2007. **40**(3): p. 191-285.
155. Svergun, D. and M. Koch, *Small-angle scattering studies of biological macromolecules in solution*. Rep Prog Phys, 2003. **66**: p. 1735-1782.
156. Hura, G.L., et al., *Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)*. Nat Methods, 2009. **6**(8): p. 606-614.
157. Toft, K., et al., *High-throughput small angle X-ray scattering from proteins in solution using a microfluidic front-end*. Anal Chem, 2008. **80**(10): p. 3648-3654.
158. Madl, T., F. Gabel, and M. Sattler, *NMR and small-angle scattering-based structural analysis of protein complexes in solution*. J Struct Biol, 2010: p. 1-11.
159. Levitt, M., *Nature of the protein universe*. Proc Natl Acad Sci U S A, 2009. **106**(27): p. 11079-11084.
160. Petoukhov, M.V. and D.I. Svergun, *Global rigid body modeling of macromolecular complexes against small-angle scattering data*. Biophys J, 2005. **89**(2): p. 1237-1250.
161. Bernadó, P., et al., *Structural characterization of flexible proteins using small-angle X-ray scattering*. J Am Chem Soc, 2007. **129**(17): p. 5656-5664.
162. Pelikan, M., G.L. Hura, and M. Hammel, *Structure and flexibility within proteins as identified through small angle X-ray scattering*. Gen Physiol Biophys, 2009. **28**(2): p. 174-189.
163. Stovgaard, K., et al., *Calculation of accurate small angle X-ray scattering curves from coarse-grained protein models*. BMC Bioinformatics, 2010. **11**: p. 429.
164. Lilyestrom, W., et al., *Structural and biophysical studies of human PARP-1 in complex with damaged DNA*. J Mol Biol, 2010. **395**(5): p. 983-994.
165. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis*. J Appl Crystallogr, 2003. **36**(5): p. 1277-1282.
166. Svergun, D.I., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria*. J Appl Crystallogr, 1992. **25**(4): p. 495-503.
167. Svergun, D., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing*. Biophys J, 1999. **76**: p. 2879-2886.
168. Kozin, M.B. and D.I. Svergun, *Automated matching of high- and low-resolution structural models*. J Appl Crystallogr, 2001. **34**(1): p. 33-41.
169. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering*. J Appl Crystallogr, 2003. **36**(3): p. 8898.
170. Wriggers, W., *Using Situs for the integration of multi-resolution structures*. Biophys Rev, 2010. **2**(1): p. 21-27.
171. Harder, T., et al., *Beyond rotamers: a generative, probabilistic model of side chains in proteins*. BMC Bioinformatics, 2010. **11**: p. 306.
172. Hartmann, C., I. Antes, and T. Lengauer, *IRECS: A new algorithm for the selection of most probable ensembles of side-chain conformations in protein models*. Protein Sci, 2007. **16**: p. 1294-1307.
173. Bishop, C.M., *Pattern recognition and machine learning* 2006, Cambridge: Springer Verlag.

174. Mardia, K.V., C.C. Taylor, and G.K. Subramaniam, *Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data*. Biometrics, 2007. **63**(2): p. 505-512.
175. Boomsma, W., et al., *A generative, probabilistic model of local protein structure*. Proc Natl Acad Sci U S A, 2008. **105**(26): p. 8932-8937.
176. Hamelryck, T., et al., *Sampling realistic protein conformations using local structural bias*. PLoS Comput Biol, 2006. **2**(9): p. 1121-1133.
177. Kent, J.T., *The Fisher-Bingham distribution on the sphere*. J Roy Stat Soc B Met, 1982. **44**: p. 71-80.
178. Wang, G. and R.L. Dunbrack, *PISCES: a protein sequence culling server*. Bioinformatics, 2003. **19**: p. 1589-1591.
179. Word, J.M., et al., *Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation*. J Mol Biol, 1999. **285**: p. 1735-1747.
180. Paluszewski, M. and T. Hamelryck, *Mocapy++ - A toolkit for inference and learning in dynamic Bayesian networks*. BMC Bioinformatics, 2010. **11**: p. 126.
181. Nielsen, S., *The stochastic EM algorithm: Estimation and asymptotic results*. Bernoulli, 2000. **6**: p. 457-489.
182. Borg, M., et al., *A probabilistic approach to protein structure prediction: PHAISTOS in CASP9*, in *LASR2009 - Statistical tools for challenges in bioinformatics2009*, Leeds University Press, Leeds, UK. p. 65-70.
183. Ferkinghoff-Borg, J., *Optimized Monte Carlo analysis for generalized ensembles*. Euro Phys J B, 2002. **29**(3): p. 481-484.
184. Svergun, D., C. Barberato, and M.H.J. Koch, *CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates*. J Appl Crystallogr, 1995. **28**: p. 768-773.
185. Schneidman-Duhovny, D., M. Hammel, and A. Sali, *FoXS: a web server for rapid computation and fitting of SAXS profiles*. Nucleic Acids Res, 2010. **38**: p. 540-544.
186. Røgen, P. and B. Fain, *Automatic classification of protein structure by using Gauss integrals*. Proc Natl Acad Sci U S A, 2003. **100**(1): p. 119.
187. Lloyd, S.P., *Least Squares Quantization in PCM*. IEEE Trans Inf Theory, 1982. **28**: p. 129-137.
188. Park, H.-S. and C.-H. Jun, *A simple and fast algorithm for K-medoids clustering*. Expert Syst Appl, 2009. **36**(2, Part 2): p. 3336-3341.
189. Rieping, W., M. Habeck, and M. Nilges, *Inferential structure determination*. Science, 2005. **309**: p. 303-306.
190. Habeck, M., W. Rieping, and M. Nilges, *Weighting of experimental evidence in macromolecular structure determination*. Proc Natl Acad Sci U S A, 2006. **103**(6): p. 1756-1761.
191. Förster, F., et al., *Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies*. J Mol Biol, 2008. **382**: p. 1089-1106.
192. Wu, Y., et al., *Folding of small helical proteins assisted by small-angle X-ray scattering profiles*. Structure, 2005. **13**(11): p. 1587-1597.
193. Grishae, A., et al., *Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data*. J Am Chem Soc, 2005. **127**(47): p. 16621-8.

194. Hesselbo, B. and R.B. Stinchcombe, *Monte Carlo simulation and global optimization without parameters*. Phys Rev Lett, 1995. **74**(12): p. 2151-2155.
195. Berman, H.M., et al., *The protein data bank*. Nucleic Acids Res, 2000. **28**: p. 235-242.
196. Sonan, G., et al., *The linker region plays a key role in the adaptation to cold of the cellulase from an Antarctic bacterium*. Biochem J, 2007. **407**: p. 293-302.
197. Svergun, D.I., M.V. Petoukhov, and M.H.J. Koch, *Determination of domain structure of proteins from X-ray solution scattering*. Biophys J, 2001. **80**(6): p. 2946-2953.
198. Amé, J.C., C. Spenlehauer, and G. de Murcia, *The PARP superfamily*. Bioessays, 2004. **26**(8): p. 882-893.
199. Tao, Z., et al., *Domain C of Human Poly (ADP-ribose) Polymerase-1 Is Important for Enzyme Activity and Contains a Novel Zinc-Ribbon Motif*. Biochemistry, 2008. **47**(21): p. 5804-5813.
200. Masson, M., et al., *XRCC1 is specifically associated with poly (ADP-ribose) polymerase and negatively regulates its activity following DNA damage*. Mol Cell Biol, 1998. **18**(6): p. 3563.
201. Bouchard, V.J., M. Rouleau, and G.G. Poirier, *PARP-1, a determinant of cell survival in response to DNA damage*. Exp Hematol, 2003. **31**(6): p. 446-54.
202. Jacobson, E.L., et al., *Poly(ADP-ribose) metabolism in ultraviolet irradiated human fibroblasts*. J Biol Chem, 1983. **258**(1): p. 103-7.
203. Oleinick, N.L. and H.H. Evans, *Poly(ADP-ribose) and the response of cells to ionizing radiation*. Radiat Res, 1985. **101**(1): p. 29-46.
204. Berger, N.A., *Poly(ADP-ribose) in the cellular response to DNA damage*. Radiat Res, 1985. **101**(1): p. 4-15.
205. McCurry, L.S. and M.K. Jacobson, *Poly(ADP-ribose) synthesis following DNA damage in cells heterozygous or homozygous for the xeroderma pigmentosum genotype*. J Biol Chem, 1981. **256**(2): p. 551-3.
206. Nolan, N.L., et al., *Characterization of poly(ADP-ribose)--histone H1 complex formation in purified polynucleosomes and chromatin*. Eur J Biochem, 1980. **113**(1): p. 15-25.
207. Petrucco, S. and R. Percudani, *Structural recognition of DNA by poly(ADP-ribose)polymerase-like zinc finger families*. Febs J, 2008. **275**(5): p. 883-93.
208. Gradwohl, G., et al., *The second zinc-finger domain of poly(ADP-ribose) polymerase determines specificity for single-stranded breaks in DNA*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(8): p. 2990-2994.
209. de Murcia, G., et al., *Structure and function of poly(ADP-ribose) polymerase*. Mol Cell Biochem, 1994. **138**(1-2): p. 15-24.
210. Lonskaya, I., et al., *Regulation of poly(ADP-ribose) polymerase-1 by DNA structure-specific binding*. J Biol Chem, 2005. **280**(17): p. 17076-83.
211. Pion, E., et al., *DNA-induced dimerization of poly(ADP-ribose) polymerase-1 triggers its activation*. Biochemistry, 2005. **44**(44): p. 14670-81.
212. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing*. Biophys J, 1999. **76**(6): p. 2879-86.
213. Kozin, M.B. and D.I. Svergun, *Automated matching of high- and low-resolution structural models*. Journal of Applied Crystallography, 2001. **34**(1): p. 33-41.
214. Koch, M.H., P. Vachette, and D.I. Svergun, *Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution*. Q Rev Biophys, 2003. **36**(2): p. 147-227.

215. Petoukhov, M.V. and D.I. Svergun, *Global rigid body modeling of macromolecular complexes against small-angle scattering data*. Biophys J, 2005. **89**(2): p. 1237-50.
216. Lipfert, J. and S. Doniach, *Small-angle X-ray scattering from RNA, proteins, and protein complexes*. Annu Rev Biophys Biomol Struct, 2007. **36**: p. 307-27.
217. Konarev, P.V., et al., *ATSAS 2.1, a program package for small-angle scattering data analysis*. Journal of Applied Crystallography, 2006. **39**(2): p. 277-286.
218. Dantzer, F., et al., *Poly(ADP-ribose) polymerase-1 activation during DNA damage and repair*. Methods Enzymol, 2006. **409**: p. 493-510.
219. Bai, Y., et al., *Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution*. Proc Natl Acad Sci U S A, 2005. **102**(4): p. 1035-40.
220. Petrucco, S., *Sensing DNA damage by PARP-like fingers*. Nucl. Acids Res., 2003. **31**(23): p. 6689-6699.
221. Pion, E., et al., *Poly(ADP-ribose) polymerase-1 dimerizes at a 5' recessed DNA end in vitro: a fluorescence study*. Biochemistry, 2003. **42**(42): p. 12409-17.
222. Kamashev, D., et al., *HU binds and folds single-stranded DNA*. Nucl. Acids Res., 2007: p. gkm667.
223. Kamashev, D. and J. Rouviere-Yaniv, *The histone-like protein HU binds specifically to DNA recombination and repair intermediates*. Embo J, 2000. **19**(23): p. 6527-35.
224. Nottbohm, A.C., et al., *A colorimetric substrate for poly(ADP-ribose) polymerase-1, VPARP, and tankyrase-1*. Angew Chem Int Ed Engl, 2007. **46**(12): p. 2066-9.
225. Kickhoefer, V.A., et al., *The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase*. J Cell Biol, 1999. **146**(5): p. 917-28.
226. Prilusky, J., et al., *FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded*. Bioinformatics, 2005. **21**(16): p. 3435-8.
227. Gagnon, S.N. and S. Desnoyers, *Single amino acid substitution enhances bacterial expression of PARP-4D214A*. Mol Cell Biochem, 2003. **243**(1-2): p. 15-22.
228. Berger, N.A., *Poly(ADP-ribose) in the cellular response to DNA damage*. 1985(0033-7587 (Print)).
229. Virag, L. and C. Szabo, *The therapeutic potential of poly(ADP-ribose) polymerase inhibitors*. Pharmacol Rev, 2002. **54**(3): p. 375-429.
230. Lewis, C. and J.A. Low, *Clinical poly(ADP-ribose) polymerase inhibitors for the treatment of cancer*. Curr Opin Investig Drugs, 2007. **8**(12): p. 1051-6.
231. Lilyestrom, W., et al., *Structural and biophysical studies of human PARP-1 in complex with damaged DNA*. J Mol Biol, 2009. **395**(5): p. 983-94.
232. Curtis, J.E., et al., *SASSIE: A program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints*. Computer Physics Communications 2012. **183**: p. 382-389.
233. Demeler, B. and K.E. van Holde, *Sedimentation velocity analysis of highly heterogeneous systems*. Anal Biochem, 2004. **335**(2): p. 279-88.
234. Koch, M.H.J., P. Vachette, and D.I. Svergun, *Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution*. Q Rev Biophys, 2003. **36**(2): p. 147-227.
235. Glinka, C.J., et al., *The 30 m Small-Angle Neutron Scattering Instruments at the National Institute of Standards and Technology*. Journal of Applied Crystallography, 1998. **31**(3): p. 430-445.

236. Cioni, P. and G.B. Strambini, *Effect of heavy water on protein flexibility*. Biophys J, 2002. **82**(6): p. 3246-53.
237. Kratky, O. and O. Glatter, *Small angle X-ray scattering* 1982: New York: Academic Press.
238. Kline, S.R., *Reduction and analysis of SANS and USANS data using IGOR Pro*. J. Appl. Cryst., 2006. **39**: p. 895-900.
239. Eswar, N., et al., *Comparative protein structure modeling using MODELLER*. Curr Protoc Protein Sci, 2007. **Chapter 2**: p. Unit 2 9.
240. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-802.
241. Svergun Di Fau - Richard, S., et al., *Protein hydration in solution: experimental observation by x-ray and neutron scattering*. 1998(0027-8424 (Print)).
242. Svergun, D.I., et al., *Protein hydration in solution: experimental observation by x-ray and neutron scattering*. Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2267-72.
243. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome*. Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
244. Oliver, A.W., et al., *Crystal structure of the catalytic fragment of murine poly(ADP-ribose) polymerase-2*. Nucleic Acids Res, 2004. **32**(2): p. 456-64.
245. Tande, B.M.a.W., N. J., *Viscosimetric, Hydrodynamic, and Conformational Properties of Dendrimers and Dendrons*. Macromolecules, 2001. **34**(24): p. 8580-8585.
246. Shiraki, K., et al., *Biophysical effect of amino acids on the prevention of protein aggregation*. J Biochem, 2002. **132**(4): p. 591-5.
247. Jacques, D.A. and J. Trehwella, *Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls*. Protein Sci, 2010. **19**(4): p. 642-57.
248. Pelikan, M., G.L. Hura, and M. Hammel, *Structure and flexibility within proteins as identified through small angle X-ray scattering*. Gen Physiol Biophys, 2009. **28**(2): p. 174-89.
249. Langelier, M.F., et al., *The Zn3 domain of human poly(ADP-ribose) polymerase-1 (PARP-1) functions in both DNA-dependent poly(ADP-ribose) synthesis activity and chromatin compaction*. 2010(1083-351X (Electronic)).
250. Nusinow, D.A., et al., *Poly(ADP-ribose) polymerase 1 is inhibited by a histone H2A variant, MacroH2A, and contributes to silencing of the inactive X chromosome*. J Biol Chem, 2007. **282**(17): p. 12851-9.
251. Ikejima, M., et al., *The zinc fingers of human poly(ADP-ribose) polymerase are differentially required for the recognition of DNA breaks and nicks and the consequent enzyme activation. Other structures recognize intact DNA*. J Biol Chem, 1990. **265**(35): p. 21907-13.
252. A. E. Whitten, S.C.a.J.T., *MULCh: modules for the analysis of small-angle neutron contrast variation data from biomolecular assemblies*. J. Appl. Cryst., 2008. **41**: p. 222-226.
253. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 Å resolution*. Nature, 1997. **389**: p. 251-259.
254. Hansen, J.C., *Conformational dynamics of the chromatin fiber in solution: Determinants, Mechanisms, and Functions*. Annu Rev Biophys Biomol Struct, 2002. **31**: p. 361-92.
255. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.

256. Jenuwein, T. and C.D. Allis, *Translating the histone code*. Science, 2001. **293**(5532): p. 1074-80.
257. Andrews, A.J. and K. Luger, *Histone Modifications: Chemistry and Structural Consequences*. Wiley Encyclopedia of Chemical Biology, 2009. **1**: p. 275-284.
258. Marmorstein, R. and R.C. Trievel, *Histone modifying enzymes: structures, mechanisms, and specificities*. Biochim Biophys Acta, 2009. **1789**(1): p. 58-68.
259. Taverna, S.D., et al., *How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers*. Nature Structural & Molecular Biology, 2007. **14**(11): p. 1025-40.
260. Lu, X., et al., *The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure*. Nature Structural & Molecular Biology, 2008. **15**(10): p. 1122-4.
261. Shogren-Knaak, M., et al., *Histone H4-K16 acetylation controls chromatin structure and protein interactions*. Science, 2006. **311**(5762): p. 844-7.
262. Robinson, P.J., et al., *30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction*. J Mol Biol, 2008. **381**(4): p. 816-25.
263. Xu, F., K. Zhang, and M. Grunstein, *Acetylation in histone H3 globular domain regulates gene expression in yeast*. Cell, 2005. **121**(3): p. 375-85.
264. Williams, S.K., D. Truong, and J.K. Tyler, *Acetylation in the globular core of histone H3 on lysine-56 promotes chromatin disassembly during transcriptional activation*. Proc Natl Acad Sci U S A, 2008. **105**(26): p. 9000-5.
265. Chen, C.C., et al., *Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair*. Cell, 2008. **134**(2): p. 231-43.
266. Tjeertes, J.V., K.M. Miller, and S.P. Jackson, *Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells*. EMBO J, 2009. **28**(13): p. 1878-89.
267. Kaplan, T., et al., *Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast*. PLoS Genet, 2008. **4**(11): p. e1000270.
268. Masumoto, H., et al., *A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response*. Nature, 2005. **436**(7048): p. 294-8.
269. Das, C., et al., *CBP/p300-mediated acetylation of histone H3 on lysine 56*. Nature, 2009.
270. Ferreira, H., et al., *Histone tails and the H3 alphaN helix regulate nucleosome mobility and stability*. Mol Cell Biol, 2007. **27**(11): p. 4037-48.
271. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N(epsilon)-acetyllysine in recombinant proteins*. Nat Chem Biol, 2008. **4**(4): p. 232-4.
272. Neumann, H., et al., *A method for genetically installing site-specific acetylation in recombinant histones defines the effects of h3 k56 acetylation*. Mol Cell, 2009. **36**(1): p. 153-63.
273. Schwarz, P.M. and J.C. Hansen, *Formation and stability of higher order chromatin structures. Contributions of the histone octamer*. J Biol Chem, 1994. **269**(23): p. 16284-9.
274. Georgel, P., et al., *Binding of the RNA polymerase I transcription complex to its promoter can modify positioning of downstream nucleosomes assembled in vitro*. J Biol Chem, 1993. **268**(3): p. 1947-54.
275. Hansen, J.C., K.E. van Holde, and D. Lohr, *The mechanism of nucleosome assembly onto oligomers of the sea urchin 5 S DNA positioning sequence*. J Biol Chem, 1991. **266**(7): p. 4276-82.

276. Hansen, J.C. and D. Lohr, *Assembly and structural properties of subsaturated chromatin arrays*. J Biol Chem, 1993. **268**(8): p. 5840-8.
277. Brunger, A.T., *Version 1.2 of the Crystallography and NMR system*. Nat Protoc, 2007. **2**(11): p. 2728-33.
278. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
279. Lu, X., et al., *In vitro chromatin self-association and its relevance to genome architecture*. Biochem Cell Biol, 2006. **84**(4): p. 411-7.
280. Gordon, F., K. Luger, and J.C. Hansen, *The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays*. J Biol Chem, 2005. **280**(40): p. 33701-6.
281. Suto, R.K., et al., *Crystal Structures of Nucleosome Core Particles in Complex with Minor Groove DNA-binding Ligands*. J Mol Biol, 2003. **326**(2): p. 371-80.
282. Schwarz, P.M., et al., *Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains*. Biochemistry, 1996. **35**(13): p. 4009-15.
283. Horn, P.J., et al., *The SIN domain of the histone octamer is essential for intramolecular folding of nucleosomal arrays*. Nat Struct Biol, 2002. **9**(3): p. 167-71.
284. Abbott, D.W., et al., *Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation*. J Biol Chem, 2001. **276**(45): p. 41945-9.
285. Fan, J.Y., et al., *The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states.[see comment][erratum appears in Nat Struct Biol 2002 Apr;9(4):316]*. Nature Structural Biology, 2002. **9**(3): p. 172-6.
286. Lee, D.Y., et al., *A positive role for histone acetylation in transcription factor access to nucleosomal DNA*. Cell, 1993. **72**(1): p. 73-84.
287. Bauer, W.R., et al., *Nucleosome structural changes due to acetylation*. J Mol Biol, 1994. **236**(3): p. 685-90.
288. Polach, K.J., P.T. Lowary, and J. Widom, *Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes*. J Mol Biol, 2000. **298**(2): p. 211-23.
289. Widlund, H.R., et al., *DNA sequence-dependent contributions of core histone tails to nucleosome stability: differential effects of acetylation and proteolytic tail removal*. Biochemistry, 2000. **39**(13): p. 3835-41.
290. Manohar, M., et al., *Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding*. J Biol Chem, 2009. **284**(35): p. 23312-21.
291. Recht, J., et al., *Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis*. Proc Natl Acad Sci U S A, 2006. **103**(18): p. 6988-93.
292. Tsubota, T., et al., *Histone H3-K56 acetylation is catalyzed by histone chaperone-dependent complexes*. Mol Cell, 2007. **25**(5): p. 703-12.
293. Bruno, M., et al., *Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities*. Mol Cell, 2003. **12**(6): p. 1599-606.
294. Vicent, G.P., et al., *DNA instructed displacement of histones H2A and H2B at an inducible promoter*. Mol Cell, 2004. **16**(3): p. 439-52.

Appendix

Structural characterization of H3K56Q nucleosomes and nucleosomal arrays

Appendix Overview

This section of this manuscript represents an auxiliary study for in which I was involved. I contributed to this work by performing Molecular Replacement on the Histone H3 K56E containing nucleosome. The problems and issues pertaining to the treatment and interpretation of x-ray diffraction data served as a fantastic learning platform for this author. Additionally, I was able to gain experience with powerful structural biology software and gain computer skills that will benefit me for the rest of his life. Without this training opportunity the author would have found completing the structural portions of chapters 4 and 5 impossible. Additionally, it was I who dealt primarily with the Protein Data Bank (PDB) representatives for the deposition of the coordinate files, electron density maps and all other associated information pertaining to the crystal structures of histone H3 K56E and histone H3 K56Q containing nucleosomes.

Summary

The posttranslational modification of histones is a key mechanism for the modulation of DNA accessibility. Acetylated lysine 56 in histone H3 is associated with nucleosome assembly during replication and DNA repair, and is thus likely to predominate in regions of chromatin containing nucleosome free regions. Here we show by x-ray crystallography that mutation of H3 lysine 56 to glutamine (to mimic acetylation) or glutamate (to cause a charge reversal) has no detectable effects on the structure of the nucleosome. At the level of higher order chromatin structure, the K to Q substitution has no effect on the folding of model nucleosomal arrays in cis, regardless of the degree of nucleosome density. In contrast, defects in array-array interactions in trans ('oligomerization') are selectively

observed for mutant H3 lysine 56 arrays that contain nucleosome free regions. Our data suggests that H3K56 acetylation is one of the molecular mechanisms employed to keep chromatin with nucleosome free regions accessible to the DNA replication and repair machinery.

Introduction

Nucleosomes, the basic building blocks of chromatin, are formed by coiling 147 base pairs of DNA around a protein core that consists of two copies each of histones H2A, H2B, H3 and H4 [253]. Hundreds of thousands of nucleosomes arrayed on chromosomal DNA undergo hierarchical condensation steps to achieve the degree of compaction that is necessary to fit the entire eukaryotic genome into the confines of the nucleus [254]. Properties that regulate the degree of compaction of nucleosomes and chromatin (e.g., histone saturation levels, histone variants, and post-translational modifications) will either locally or globally affect DNA accessibility to permit access to the genome.

Posttranslational modifications of histones have emerged as a key mechanism to regulate important biological processes such as transcription, DNA repair, and replication (reviewed for example in [255, 256]). Numerous side chains in the histone tails, and an increasing number of amino acids in the structured regions of the histones, are the targets of tightly regulated and highly specific activities that add or remove chemical modifications to specific locations in chromatin in response to biological cues (reviewed in [255, 257, 258]). Modern techniques identify new post-translational modifications at a rapid rate; however, our understanding of the mechanisms by which the chemical modification of selected histone residues affects chromatin biology has lagged behind. In many cases, specifically modified histone tails in turn recruit specific activities required for the required task (e.g. DNA repair, transcription, etc.; [259]). Recent structural work has confirmed the notion that the structure of the nucleosome itself is not greatly affected by histone modifications [260]. However, modification of certain residues has pronounced effects on the ability of nucleosomal array to fold and compact into higher order structures of increasing complexity [260-262].

Recently, the acetylation of H3K56 has received much attention due to its implied biological roles in transcription, DNA repair, and in maintaining genomic stability [263-266]. The modification is added onto non-nucleosomal H3 by the HAT Rtt109 (in yeast) or p300 (in metazoans) and is subsequently incorporated into nucleosomes. H3K56ac is a marker for newly synthesized histones during replication ([267] and references therein), and is also implicated in creating a favorable chromatin environment for DNA repair [268, 269]. Additionally, it plays a role in chromatin disassembly during transcriptional activation [263, 264]. Because of the location of this residue in the structured region of H3 near the DNA at its entry- and exit point [253] it has been speculated that acetylation may destabilize the nucleosome sufficiently to account for some of the observed biological effects [270].

A recent exciting technical development now allows genetic encoding of N(epsilon)-acetyllysine into recombinant proteins in specific positions [271]. This has enabled a first analysis of nucleosomes and nucleosomal arrays reconstituted with histone H3 specifically acetylated at K56 [272]. These studies showed that H3K56ac does not affect salt-dependent nucleosome stability, but that moderately increased 'breathing' of the DNA ends can be observed in H3K56ac nucleosomes. It was further shown that there is no effect of this modification on the salt-dependent compaction of saturated nucleosomal arrays with and without linker histone H5. However, a long saturated array of nucleosomes with linker histone does not mimic the natural chromatin configurations in which H3K56ac is found (see above). Due to the preponderance of H3K56ac in regions of active transcription and near sites of replication- and repair-coupled DNA assembly, this modification should also be studied in the context of subsaturated chromatin depleted of nucleosomes and linker histone, i.e., nucleosomal arrays containing nucleosome-free regions.

Here we present two crystal structures of nucleosomes in which H3K56 has been substituted with either glutamine to mimic acetylation, or with glutamic acid to introduce a charge reversal at this location. Our data indicate that the structure of the nucleosome remains unaffected by these changes. We also analyze the folding and oligomerization properties of subsaturated and saturated nucleosomal arrays bearing K56Q and find that this substitution negatively affects the ability of the arrays to oligomerize when they are subsaturated and contain nucleosome-free "gaps" in the arrays. No effects of K56Q were

observed at the level of local array folding. Our data suggests that the acetylation of H3K56 results in a more globally open and accessible chromatin structure in regions of the genome depleted of nucleosomes.

Materials and Methods

Proteins and DNA

Histone expression and purification was performed as described previously [143]. The 147 bp palindromic α -sat DNA was purified as described [143]. The 208-12 5S rDNA repeat used to prepare model nucleosomal arrays was purified following published procedures [273, 274].

Reconstitution of nucleosomes and nucleosomal arrays

Nucleosomes were reconstituted onto palindromic 146 bp DNA fragment derived from α -satellite DNA as previously described [143, 253]. Briefly, equal molar ratios of histone octamers containing H3K56Q or H3K56E were mixed with 147 bp α -sat DNA in TE buffer (10mM Tris-HCl, 0.25 mM EDTA, pH 7.5) containing 2.0 M KCl₂. and dialyzed using salt gradient dialysis into TE buffer (0 M KCl₂). Nucleosomes were heat shifted at 37°C for 1 hour to uniformly position the octamer on the 147 bp DNA template. The nucleosomes were purified from excess DNA and unbound protein using a Prep Cell Model 491 purification system (Bio-Rad) and analyzed by native – PAGE [143].

Nucleosomal arrays were reconstituted onto 208-12-5S rDNA as described [275]. Briefly, equimolar ratios of histone octamers containing H2A, H2B, H4 and H3 or H3K56Q were mixed with the 208-12 DNA template in TE buffer (10 mM Tris-HCl, 0.25 mM EDTA, pH 7.8) containing 2.0 M NaCl₂, followed by step-wise salt gradient dialysis to low salt TEN buffer (2.5 mM NaCl₂-TE) [275, 276].

Nucleosome crystallization

Nucleosomes containing H3K56Q and H3K56E were crystallized by using salting in vapor diffusion at nucleosome concentrations ranging from 8-10 mg/ml and solution conditions of 36 mM KCl,

40 mM MnCl₂, 5 mM K-cacodylate and 36 mM KCl, 42 mM MnCl₂, 5 mM K-cacodylate. The crystals were soaked in 24% 2-methyl, 2,4-pentanediol (MPD), 5% trehalose, 40.0 mM KCl, 37.0 mM MnCl₂ 5.0 mM K-Cacodylate, pH 6.0 [253]. X-ray data was collected at the Advanced Light Source (beam line 4.2.2). The data was processed with Denzo and Scalepack. PDB entry 1KX5 was used as a search model for molecular replacement. Molecular replacement and further refinement was done with CNS [277]. Coot was used for model building [278].

Analytical ultracentrifugation

Sedimentation velocity studies were carried out using a Beckman XL-A analytical ultracentrifuge using absorbance optics. Samples were mixed to a final A₂₆₀ of 0.6 - 0.8 and equilibrated at 200 C. for one hour prior to sedimentation. Nucleosomal arrays were sedimented at 18-25,000 rpm with radial increments of 0.001cm. \bar{V} and ρ were calculated using Ultrascan v9.4 for windows.

Folding and oligomerization of nucleosomal arrays

To assay folding, nucleosomal arrays were diluted with TEN buffer to a final concentration of 2.0 mM MgCl₂ and a final A₂₆₀ of 0.6 – 0.8 and subjected to sedimentation velocity. Boundaries was analyzed using the improved method of van Holde and Weischet [233] to obtain the integral distribution of sedimentation coefficients, G(s), using UltraScan v9.4 for windows. The average sedimentation coefficient (s_{mid}) was defined as the sedimentation coefficient at the boundary midpoint (boundary fraction = 0.5).

To assay oligomerization, differential centrifugation was used as previously described [279, 280]. Briefly, nucleosome arrays were diluted to an A₂₆₀ = 1.2 with TEN buffer. Arrays were mixed with MgCl₂-TEN buffer, incubated for five minutes at room temperature and then centrifuged in a bench-top microfuge at 13,000 RPM (~16,000 x g) for 5 min. The A₂₆₀ of the supernatant was then determined in a Beckman DU 800 Spectrophotometer. Data were expressed as a percentage of the total sample that

remained in the supernatant as a function of MgCl₂. The Mg₅₀ is defined as the MgCl₂ concentration at which the sample was 50% oligomerized [279, 280].

Results:

The crystal structures of nucleosomes containing H3K56Q and H3K56E.

We used site directed mutagenesis of H3 to mimic acetylation (H3K56Q), and to introduce a more extreme disturbance at this site by changing the charge from positive to negative (H3K56E). Recombinant H3 containing H356Q or H3K56E were assembled into histone octamers together with recombinant H2A, H2B, and H4. Mono-nucleosomes were reconstituted onto palindromic α -sat DNA [253]. Both mutant nucleosomes were indistinguishable from wild type unmodified nucleosomes based on heat-shifted mobility changes in EMSA assays (**Figure A.1A, B**).

We determined the crystal structure of nucleosomes reconstituted with H3K56Q and H3K56E to a resolution of 3.8 and 3.2 Å respectively, using molecular replacement. **Table A.2.1** summarizes data collection and refinement statistics for both datasets. H3K56 is located near the entry and exit point of nucleosomal DNA (**Figure A.2A, B**). The N epsilon group of H3K56 has a distance of ~ 4 Å to the nearest phosphate group in nucleosomal DNA (**Figure A.2C**). The mutation of H3K56 to Q or E had no effect on the overall structure of the nucleosome, as revealed by rmsds of < 0.5 when comparing either structure to the wild type nucleosome. The H356E side chain is clearly visible in the electron density generated in omit maps (**Figure A.2C**). The acidic side chain rotates away from the DNA and towards the solvent. The density for H3-56Q is not visible due to the relatively low resolution of this structure. However, the main conformation of the main chain of the histone and the C β atom of H356Q is not altered compared to the wild type structures. Importantly, DNA conformation was also unchanged in both particles compared to wild type nucleosomes. This was not unexpected, since the distance between H3K56 and the DNA phosphodiester

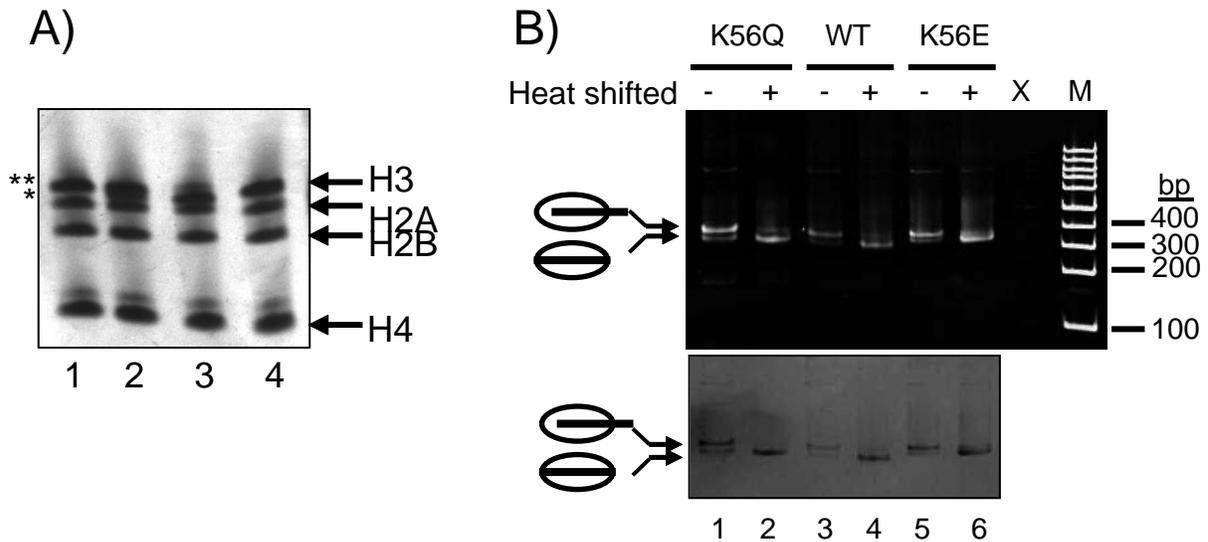


Figure A.1. Nucleosomes reconstituted with H3K56Q and K56E. A) Mutation of H3K56 to Q or E results in reduced electrophoretic mobility on triton-urea gels. 4 μ g of refolded histone octamers containing mutations to histone H3K56Q (lane 1), H3K56Q with WT H3 (lane 2), WT (lane 3) and H3K56E (lane 4) were loaded and electrophoresed on an AU-PAGE gel stained with coomassie blue. () indicates wild type H3 and (■■) indicates H3K56Q and H3K56E mobility. B) Reconstituted nucleosomes containing H3- K56 mutations have a similar ability to heat shift. Nucleosomes containing H3 K56Q (lanes 1 and 2), WT H3 (lane 3 and 4) and H3 K56E (lane 5 and 6) reconstituted with 147 bp α -Sat DNA. Nucleosomes were heat shifted (+) for 1 hour at 370C (lanes 2, 4 and 6). Samples were separated on a 5 % native polyacrylamide gel and stained with ethidium bromide (top) and Imperial Protein Stain (Pierce) (bottom). M indicates 100 bp marker was used.

Table A.1: data collection and refinement statistics for nucleosomes reconstituted with H3K56Q and H3K56E.

| <i>Data Collection</i> | <i>K56Q</i> | <i>K56E</i> |
|--|---|---|
| Space Group | P2 ₁ 2 ₁ 2 ₁ | P2 ₁ 2 ₁ 2 ₁ |
| Cell dimensions | | |
| <i>a</i> , <i>b</i> , <i>c</i> (Å) | 109.54, 105.56, 180.65 | 109.51, 105.67, 180.56 |
| α, β, γ | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å) | 20 (3.78) | 20 (3.2) |
| R _{sym} or R _{merge} | 0.109 (0.358) | 0.055 (0.342) |
| I/ σ I | 6.0 (4.3) | 19.53 (3.9) |
| Completeness % | 100 (100) | 81(73.6) |
| Redundancy | 7.23 (7.41) | 2.2(2.0) |
| <i>Refinement</i> | | |
| Resolution (Å) | 3.78 | 3.2 |
| Observations | 308,860 (26,559) | 372,562 (35,641) |
| R _{work} /R _{free} | 0.2684/0.3146 | 0.2845/0.2926 |
| No. Atoms | | |
| Protein | 6012 | 5831 |
| DNA | 5980 | 5980 |
| B-factors | | |
| Protein | 87.5 | 50.1 |
| DNA | 147 | 109.6 |
| R.m.s. deviations | | |
| Protein bond lengths | 0.013 | 0.013 |
| DNA bond lengths | 0.005 | 0.005 |
| Protein bond angle | 1.711 | 1.642 |
| DNA bond angle | 0.808 | .762 |

backbone is too long for a hydrogen bond, and since there is plenty of space for the mutated amino acid to avoid charge-charge interference. Because the side chain is free to rotate, it is unlikely to cause charge-charge repulsion with the DNA. It has been previously observed that the acetylation of H3K56 results in increased breathing of DNA ends [272]. Crystal packing stabilizes the DNA in a ‘closed’ conformation [281], and thus our results are entirely consistent with this observation.

H3K56Q affects oligomerization of nucleosomal arrays that are depleted of nucleosomes.

In vitro chromatin condensation consists of two reversible salt-dependent structural transitions: folding (mediated by short-range nucleosome – nucleosome interactions within one nucleosomal array) and oligomerization (mediated by inter-array interactions) [254, 260, 261]. To investigate the effect of H3K56 modification on higher order chromatin structure, we assembled model nucleosomal arrays from 208-12 DNA and histone octamers containing the acetylation mimic, H3K56Q. The condensation behavior of wild type and mutant nucleosomal arrays via inter-array interactions was then determined as a function of MgCl₂ concentration. Because the acetylation of H3K56 is predominantly associated with chromatin assembly and nucleosome-free genomic regions, we wanted to test the effect of H3K56Q on the condensation properties of not only saturated templates, but subsaturated templates containing nucleosome-free repeats [276].

The oligomerization curves for nucleosomal arrays that had an average ~11 bound histone octamers per 208-12 DNA template are shown in **Figure A.3A**, black symbols. The degree of saturation was determined by sedimentation velocity in TEN buffer (smidpoint = 25-26.5 S; data not shown) as described [275, 276]. Under these conditions, the MgCl₂ concentration at which the sample was 50% oligomerized (Mg₅₀) of H3K56Q arrays was indistinguishable from wild type arrays. However, when the oligomerization experiment was repeated with subsaturated nucleosomal arrays containing an average of only ~8 histone octamers/template (smidpoint = 21-22 S; data not shown), the oligomerization curve and Mg₅₀ of the H3K56Q arrays were right-shifted relative to wild type (**Figure A.3A**, open symbols). A

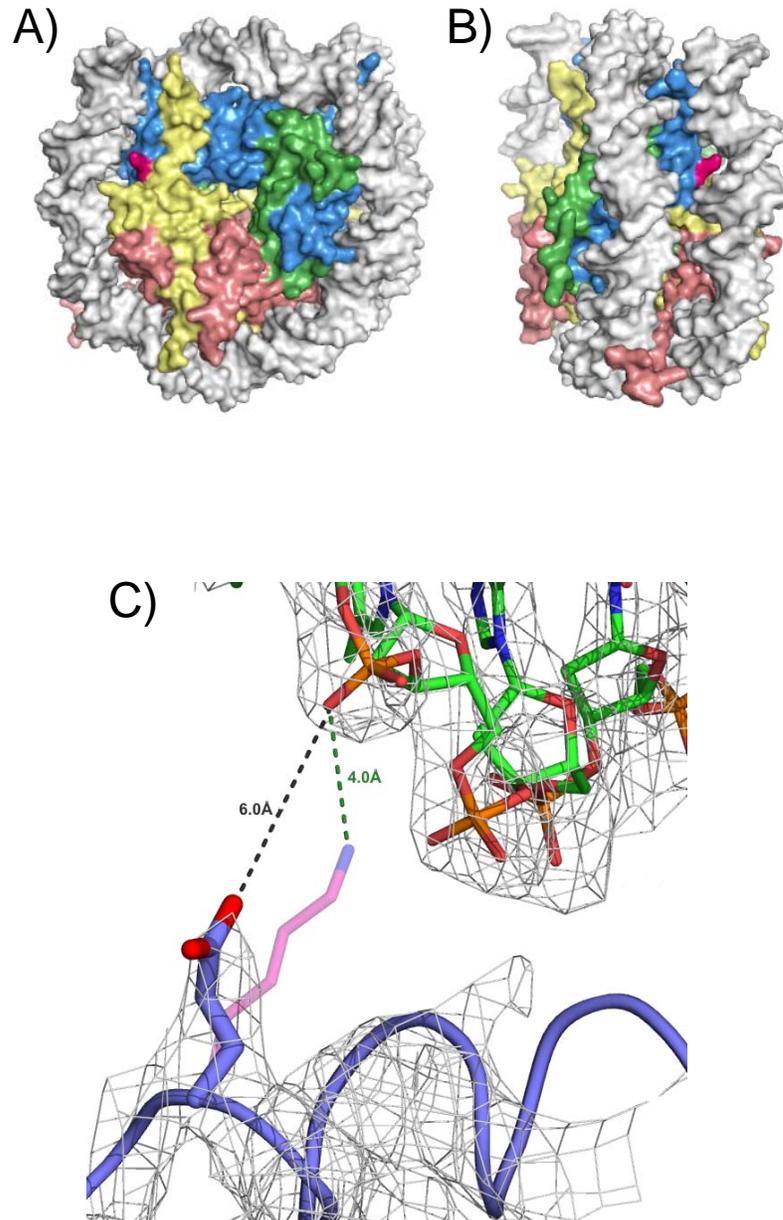


Figure A.2: Structural analysis of nucleosomes reconstituted with H3K56E. A) Location of H3K56 in the nucleosome. H3 is shown in blue, H4 in green, H2A in yellow, and H2B in red. H3K56 is indicated in magenta. The nucleosome is viewed down the superhelical axis. B) as A), but rotated around the y-axis by about 75 degrees. C) Detailed view of the structure of a nucleosome with H3K56E, superimposed onto the wild type structure. Distances between the side chains and DNA are indicated. Electron density ($2F_o - F_c$) is contoured at 1 sigma.

summary of the oligomerization data obtained over a wide range of saturation levels is shown in **Figure A.3B**.

Data are plotted as Mg50 against the number of nucleosomes as determined by sedimentation coefficient in TEN. The data points are fit with a linear regression to demonstrate the overall trend, even though the distribution of points for H3K56Q is suggestive of nonlinear behavior. The results showed two clear trends. First, the Mg50 of the wild type arrays increased linearly as the extent of saturation decreased, consistent with previous results [282]. Second H3K56Q had a major effect on oligomerization of highly to moderately subsaturated arrays. Of note, the effect of K56Q on oligomerization was greater than the effect of the loss of histone octamers per se. These data demonstrate that the H3K56Q mutation is able to significantly disrupt cooperative, inter-array interactions when the arrays contain nucleosome-free regions.

Folding of nucleosomal arrays is unaffected by the H3K56Q mutation.

We next examined the salt-dependent folding of wild type and H3K56Q nucleosomal arrays as a function of array saturation. Folding is assayed increases in the sedimentation coefficient at lower salt concentrations than cause oligomerization. S-values were identical for wild type and H3K56Q arrays in low salt TEN at all levels of saturation. Samples were exposed to 1.75 mM MgCl₂ to induce folding and subjected to sedimentation velocity in the analytical ultracentrifuge. Data were analyzed to obtain the integral distribution of sedimentation coefficients across the boundaries [233]. Figure 4A shows the data obtained for saturated arrays. Both the wild type and H3K56Q arrays folded robustly as indicated by the right-shifted sedimentation coefficient distributions that ranged as high as 55S [254]. Importantly, the wild type and H3K56Q profiles were essentially identical, indicating no effect of this modification on folding under these conditions. A graph of smidpoint against the number of nucleosomes per template firms that wild type and K56Q arrays fold identically over a wide range of nucleosome saturated levels

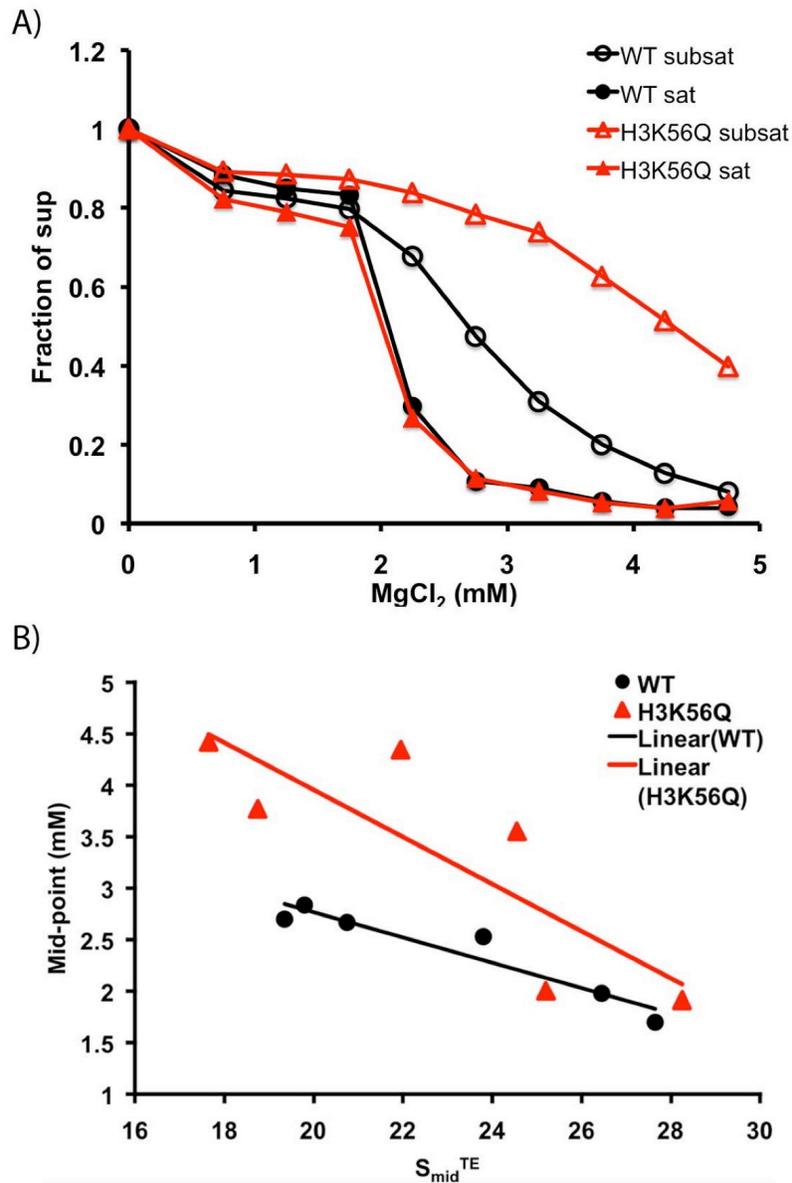


Figure A.3: H3K56Q disrupts intermolecular oligomerization of subsaturated nucleosomal arrays.

A) Intermolecular oligomerization assay. Nucleosomal arrays were incubated with varying concentrations of MgCl₂, followed by centrifugation. The fraction of array remaining in the supernatant is plotted as a function of MgCl₂ concentration. Samples of the following saturation levels, indicated by SmidTE were analyzed: black circles, filled: wild type, 26.5S; black circles, open: wild type 20.8S; red triangles, filled: H3K56, 25.2S; red triangles, open: H3K56Q, 22S. B) Effect of nucleosomal array saturation on intermolecular oligomerization. Nucleosomal arrays were reconstituted at various ratios of histone octamer to 208 bp 5S DNA repeats. Arrays were analyzed by sedimentation velocity analysis in TE buffer to determine their degree of saturation (SmidTE). MgCl₂ concentration required for half of the arrays to remain in the supernatant is plotted as a function of SmidTE. Black circles, wild type; red triangles, H3K56Q. Lines represent a linear regression through the data points.

(Fig. 4B). These data suggest that H3K56Q does not affect folding of nucleosomal arrays.

Discussion

Posttranslational modifications of histones have the potential to alter chromatin structure at many levels. Modification or amino acid substitutions of the histone tails (e.g. [260, 261, 283] or introduction of histone variants to alter the nucleosome surface (e.g. [284, 285]) can change higher order chromatin folding and oligomerization. Post-translational modifications of histones can also have moderate effects on accessibility and DNA conformation in a mono-nucleosome [286-289], and may alter the ability of histones to engage in histone-histone and histone-DNA interactions, with potential effects on nucleosome stability as was recently observed for the acetylation of H3K56 [272]. It is therefore important to characterize the effect of each histone post-translational modification at multiple structural levels and in different contexts.

The vast majority of histone modifications are located in the flexible histone tails, with little if any potential to impact the structure of the nucleosome per se. However, an increasing number of posttranslational modifications are being identified in the structured region of the histones, with several on side chains near the DNA. Thus far three from the latter category have been characterized biophysically and structurally. The dimethylation of H3K79 has no effect on either the structure nucleosomes or the condensation of nucleosomal arrays [260]; the acetylation of H3K115 and H3K122 increases the rate of thermal repositioning [290], with no implications for the structure (M.L. Dechassaand K.L., unpublished); and the acetylation of H3K56 results in subtle changes in the exposure rate of DNA ends in mononucleosomes [272].

H3K56Q mimics constitutive acetylation in that it causes reduced superhelicity of plasmid chromatin isolated from yeast cells and more rapid nuclease digestion of cellular chromatin [268, 291] and thus represents a good model for structural studies of H3K56Ac. We find that the substitution of

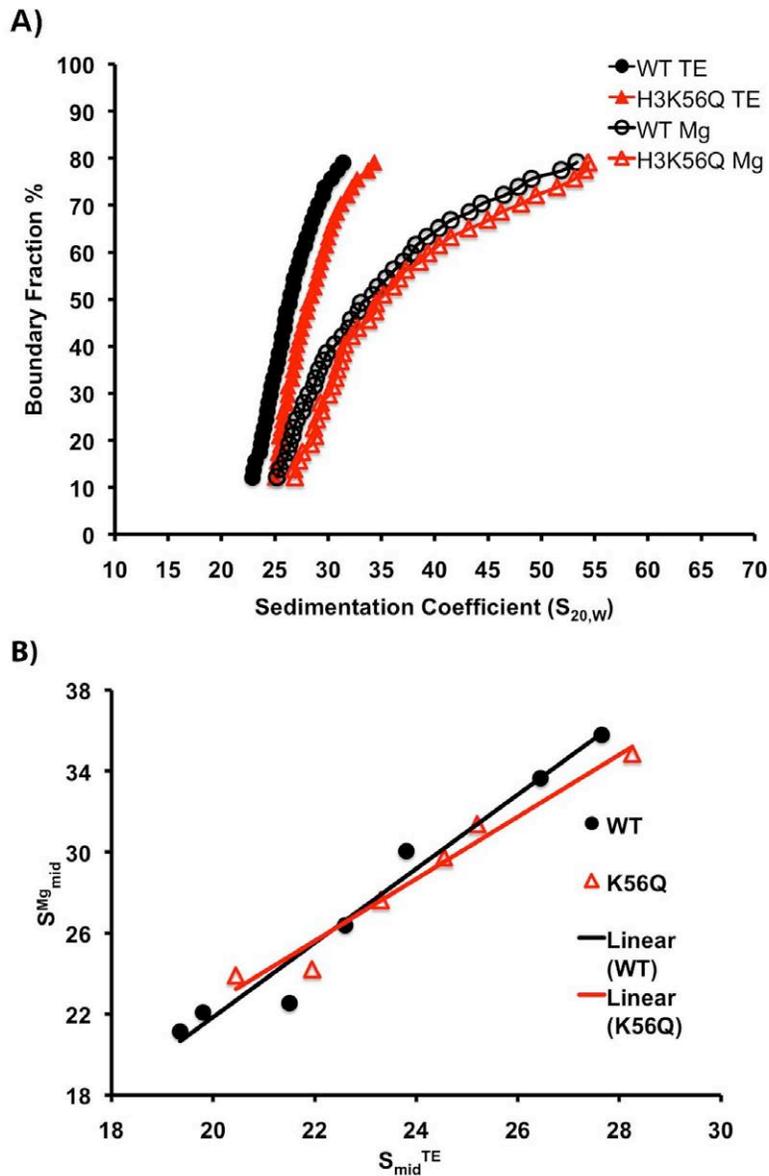


Figure A.4. H3K56Q does not affect intramolecular folding of nucleosomal arrays. A) Sedimentation velocity analysis of 208-12 nucleosomal arrays in TE (10mM Tris-HCl (pH7.4) and 0.25mM EDTA) or TE with 1.75mM MgCl₂. Integrated sedimentation coefficient distributions of nucleosomal arrays were determined by sedimentation velocity and van Holde-Weischet analysis. $S_{20,w}$ is the sedimentation coefficient corrected to water at 20 °C. Black circles: wild type arrays in TE; open circles: wild type arrays in MgCl₂. Red triangles: H3K56Q arrays in TE; open triangles, H3K56Q arrays in MgCl₂. B) Intramolecular folding of nucleosomal arrays at varying levels of nucleosome occupancy. Black circles: wild type arrays, red triangles: H3K56Q arrays. Nucleosomal arrays were reconstituted at various ratios of histone octamer to 208 bp 5S DNA repeats. Arrays were analyzed by a sedimentation velocity analysis in either TE or TE + 1.75 mM MgCl₂ buffers. The S_{mid}^{TE} and S_{mid}^{Mg} are defined as the sedimentation coefficients at the boundary fraction = 0.5 in TE and TE with 1.75mM MgCl₂, respectively. Lines represent a linear regression through the data points.

H3K56 with either Q or E (resulting in an acetylation mimic or charge reversal, respectively), results in no discernable structural changes either in the histone octamer or in the path of the DNA.

Since the canonical crystal packing of nucleosome entails base stacking of the DNA as it enters and exits the nucleosome [281], the crystal lattice selects for nucleosomes in which the DNA is in close contact with the histone octamer; and thus subtle differences in DNA dynamics, as demonstrated by Neumann and colleagues [272] cannot be detected through x-ray crystallography.

Lysine 56 acetylation is an abundant modification of newly synthesized histone H3 molecules that are incorporated during S phase and during DNA damage repair [292]. Our finding that the acetylation-mimic H3K56Q affects the oligomerization of nucleosomal arrays only in the context of subsaturated arrays containing nucleosome-free regions is consistent with this important role. The oligomerization of nucleosomal arrays reflects long-range interactions between nucleosomes that are distinct from the short-range interactions responsible for array folding [254]. The independence of the two condensation transitions is reinforced by our result that the same K56Q mutation that disrupts oligomerization of subsaturated nucleosomal arrays has no effect on array folding, irrespective of their saturation level.

At this point one can only speculate on the physical basis for the effect observed with subsaturated arrays. It seems likely that our results are related to the increased site exposure of nucleosomal DNA observed by Neumann and colleagues in H3K56Ac nucleosomal arrays [19]. We speculate that site exposure may be more pronounced in arrays with multiple long nucleosome free regions, which serves to free up more linker DNA in the subsaturated arrays. This subsequently would require more MgCl₂ to induce oligomerization, as we have observed. It is also possible that H3K56 constitutes a surface area of the nucleosome that interacts with the histone tails during oligomerization, but not folding, and that H3K56Ac is more effective at disrupting such an interaction if there are large gaps in the array. Clearly, further experiments are required to dissect the molecular basis for the unique behavior of subsaturated arrays during oligomerization.

Our findings have several important ramifications for genomic regions harboring nucleosomes with acetylated H3K56. The presence of genomic chromatin containing nucleosome free regions is linked to processes involving increased DNA accessibility, such as transcription, repair, and replication [263-266, 268, 269]. As discussed above, H3K56Ac tends to be associated with the nucleosomes in and around nucleosome free regions. The surprising observation that oligomerization of nucleosomal arrays containing nucleosome free regions is selectively affected by the K56Q mutation, with H3K56Q destabilizing inter-array interactions *in vitro*, strongly suggests that H3K56 acetylation is one of the mechanisms employed to keep chromatin with nucleosome free regions accessible at the higher order level. Both our data with H3K56Q, and those obtained previously with H3K56Ac [272], indicate that this modification does not function at the level of local chromatin fiber folding. However, we note that folding is intrinsically inhibited by nucleosome free regions [273, 274]. Furthermore, the destabilization of (H3-H4)₂ tetramer-DNA interactions by H3K56Ac (A. J. Andrews and K.L., unpublished) may also facilitate nucleosome eviction events at RNAPII promoters, subsequent to the removal of the H2A-H2B dimer by ATP-dependent nucleosome remodeling enzymes or histone chaperones [293, 294]. Regardless of the mechanism(s) involved, identification of a modification that specifically targets chromatin with nucleosome free regions is novel. Our data are consistent with a model in which the tertiary interactions of subsaturated nucleosomal arrays are regulated by acetylation of H3 K56, and disruption of such interactions by H3 K56Ac has a significant impact on DNA repair, transcription, and the stability of stalled replication forks.

Acknowledgements

We thank Teri McLain from the W. M. Keck PEPF for histones, Pamela N. Dyer and Alyson White for DNA preparations, Mark van der Woerd and Jay Nix for help with structure determination. Supported by the NIH (GM045916 to JCH, GM54096 to CLP, and GM061909 to KL.) Coordinate files for the two structures have been deposited at the protein data bank (PDB ID 3KWQ and 3KXB).

This work was completed through the collaborations of many individuals. The following list of contributors to this work is listed in order of contribution to the peer reviewed article published in the journal of *Biochim Biophys Acta*; Shinya Watanabe^{*1}, Michael Resch^{*2}, Wayne Lilyestrom², Nicholas Clark², Jeffrey C. Hansen^{2**}, Craig Peterson^{1**}, and Karolin Luger^{2, 3**}

* These authors contributed equally to this work

1-Program in Molecular Medicine, University of Massachusetts Medical School, 373 Plantation St.; Worcester, Massachusetts 01605

2-Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO 80523-1870

3-Howard Hughes Medical Institute

** Corresponding authors