Title

Advances in X-ray Scattering Interferometry and Liquid Chromatography Coupled Small Angle X-Ray Scattering Towards Exploring DNA-Protein Interactions and Bionanoengineering

Permalink

https://escholarship.org/uc/item/1x91j6pp

Author

Rosenberg, Daniel J

Publication Date

2023

Peer reviewed|Thesis/dissertation

Advances in X-ray Scattering Interferometry and Liquid Chromatography Coupled Small Angle X-Ray Scattering Towards Exploring DNA-Protein Interactions and Bionanoengineering

By

Daniel Joshua Rosenberg

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Dr. Gregory Hura, Co-Chair

Professor Markita P. Landry, Co-Chair

Professor Eva Nogales

Professor James Hurley

Spring 2023

Abstract

Advances in X-ray Scattering Interferometry and Liquid Chromatography Coupled Small Angle X-Ray Scattering Towards Exploring DNA-Protein Interactions and Bionanoengineering

by

Daniel Joshua Rosenberg

Doctor of Philosophy in Biophysics

University of California, Berkeley

Dr. Gregory Hura, Co-Chair

Professor Markita P. Landry, Co-Chair

This dissertation demonstrates the development and application of advanced X-ray scattering techniques toward the exploration of DNA-protein interactions and engineering bionanotechnology. Herein I describe the development of size exclusion chromatography-coupled small-angle X-ray scattering with in-line multi-angle light scattering (SEC-SAXS-MALS) and X-ray scattering interferometry (XSI). SEC-SAXS-MALS is a high-throughput, multimodal structural biology approach with integrated purification whereas XSI leverages the interference pattern between ordered gold nanoparticles (AuNPs) to calculate discrete inter-AuNP distances, effectively turning them into molecular rulers. I demonstrate the application of these techniques in the study of the single-strand break (SSB) DNA-damage repair meiotic recombination 11 (MRE11) nuclease activity, poly(ADP-ribose) polymerase 1 (PARP-1) damage recognition, activation, and inhibition, and the dynamics of the DNA-dependent protein kinase holoenzyme DNA-PK complex. Additionally, I show the versatility of these techniques in bionanoengineering by exploring the surface adsorbed morphology and nanosensing mechanisms of DNA-functionalized single-walled carbon nanotubes (ssDNA-SWCNTs), and the innovation of RuBisCO assemblies. Finally, I show how these techniques provide high-throughput capabilities that enable rapid disaster responses by characterizing the Nsp7, Nsp8, and Nsp12 complexes of SARS-CoV-2 involved in replication, and the exploration of rigid monoclonal antibodies to improve detection of the SARS-CoV-2 nucleocapsid.

Table of Contents

List of Figures

List of Tables

Acknowledgments

It is impossible to express in words the gratitude I feel for my family for their unwavering support (except for that one time…) over the years.

Thank you to my partner Elyse Schriber for always understanding. Knowing that you were with me and suffering just as much through our academic pursuits has given me strength when doubt may have overwise consumed me.

Thank you to my heterosexual life-mate Tony Merced for always being there when I needed someone the most and for always having a place (closet or couch or driveway) for me to sleep.

My deepest gratitude to all my mentors over the years. Most importantly Greg Hura and Michal Hammel for everything you have taught me and for your considerable patience. I am fully aware that it must not have been easy for you to employ someone as obstinate and mercurial as myself. Sorry for the extra grey hair and sleepless nights.

Thank you to all my collaborators over the years for which this work would not have been possible (or at least would have taken much much longer). Disclaimer: during the quarantine years of the 2020 SARS-CoV2 pandemic, many of these collaborators I never actually met so they might be terrible in real life.

Finally for the rest of you who have known me over the course of my education… I blame you all. The last six months of my doctoral work and most notably the writing of this dissertation have truly been an exercise in sustained suffering giving me renewed respect for the fortitude of the human spirit. For future colleagues and patrons of this work, you may assuage yourself of guilt but those who know me currently upon the writing of this dissertation, know that you have played a role in prolonging my agonies with your constant support, encouragement, and love. You all know who you are, and you owe me…

1 Introduction

1.1 X-ray Scattering at the Structurally Integrated BiologY for Life Sciences (SIBYLS) Beamline[*]

Small-angle X-ray Scattering (SAXS) is a convenient and versatile technique for studying the structure, complexation, and activity of biomolecules in solution[1]. The method utilizes X-ray radiation to observe changes in the electron density of macromolecules as they exist in solution at near biological conditions. While commercial, "tabletop" SAXS instruments are becoming more widely available, the most popular method still involves the use of synchrotron facilities like the Advanced Light Source (ALS) in Berkeley, California, due to the significantly increased flux of the X-ray source. This increased brightness consequently reduces the exposure times needed from minutes or hours down to seconds. Leveraging the critical reduction in data collection time, SAXS beamlines at synchrotrons worldwide have been able to expand upon traditional SAXS collection methods to develop new techniques with a high-throughput, multimodal design philosophy.

In 2001, the **S**tructurally **I**ntegrated **B**iolog**Y** for **L**ife **S**ciences (SIBYLS) beamline (bl12.3.1)[2] was commissioned at the ALS to be the flagship BioSAXS[1] beamline at Lawrence Berkeley National Laboratory (LBNL). This beamline specializes in technologies for the characterization of biological molecules and has been a staging ground for the development of integrated biophysical techniques including size exclusion chromatography coupled SAXS with in-line multi-angle light scattering (SEC-SAXS-MALS),[3-8] high-throughput SAXS (HT-SAXS),[9,10] and X-ray scattering interferometry (XSI).[11,12] SEC-SAXS-MALS integrates *in situ* purification with a high-throughput, multimodal design philosophy revolutionizing the reach and tempo of BioSAXS experiments. HT-SAXS utilizes robotics to automate rapid data collection with a 96-well plate being collected in under 90 minutes with less than 25 μL of sample required per well. XSI leverages the interference pattern produced by AuNP tags conjugated to DNA to act as molecular rulers in solution greatly increasing spatial resolution while reducing sample requirements. All systems are capable of automated data collection with the development of SEC-SAXS-MALS and XSI described in detail within this dissertation.

1.2 Development of Size Exclusion Coupled Small Angle X-ray Scattering with Tandem Multi Angle Light Scattering (SEC-SAXS-MALS) [*]

Size exclusion chromatography coupled small-angle X-ray scattering with in-line multi-angle light scattering (SEC-SAXS-MALS) is a high-throughput, multimodal approach to studying structural biology with integrated *in situ* purification. These integrated techniques provide immediate validation for the analogous values they provide. The molecular weight calculated from the MALS (MW_{MALS}) is analogous to the molecular weight calculated from the SAXS

---

[*] Part of this section is published as Rosenberg, D. J.; Hura, G. L.; Hammel, M. Size Exclusion Chromatography Coupled Small Angle X-Ray Scattering with Tandem Multiangle Light Scattering at the SIBYLS Beamline. In *Methods in Enzymology*; Elsevier, 2022; Vol. 677, pp 191–219.

intensity at the zero angles, I(0), (MW$_{SAXS}$). One angle from the MALS is isolated to generate quasi-elastic light scattering (QELS, a.k.a. dynamic light scattering) data providing values for the radius of hydration (Rh) analogous to the radius of gyration (Rg) calculated from the SAXS. The continued development of SEC-SAXS-MALS builds upon an expanding pool of previous work.[13-21] This dissertation, describes the design, collection, and processing of the SEC-SAXS-MALS system developed at the SIBYLS beamline in detail as well as highlighting its effectiveness at exploring DNA damage repair and bionanoengineering.

1.3 Development of High-throughput X-ray Scattering Interferometry (HT-XSI)

The development of high-throughput X-ray scattering interferometry (HT-XSI) has come about through the simultaneous development of high-throughput beamline technologies and the strategic development of DNA-AuNPs substrates. Since 2018, the SIBYLS beamline has seen several major upgrades, which complement the incorporation of HT-XSI, including a new shutterless Dectris Pilatus 2M detector and the integration of a modified Tecan EVO liquid handling robot to increase framerates and throughput respectively. The major improvement to the HT-SAXS system involves custom-designed 3D printed liquid handling needles, which effectively turn each needle into the sample cells greatly decreasing collection times and volumes (~1 sample per min with 25 μL per sample).[22,23]

The preparation of colloidal gold nanoparticles (AuNPs) has been known for over 150 years[24] and the synthesis of sub 10 nm AuNPs of tunable size and morphology is very well documented[25,26], but the functionalization of DNA on the gold surface is relatively new technology[27,28]. In 2008, Mathew-Fenn et al. were the first group to use dsDNA-AuNPs as molecular rulers via XSI[29] and went on to apply this technique to remeasuring the double helix[30]. The 0.7 and 1.4 nm AuNPs used in these studies were prepared by the Brust method[31,32] and coupled to the 3′-ends of dsDNA via sulfhydryl-gold bonds. The interference pattern generated by solution SAXS of the two AuNPs was decomposed into a linear combination of basis patterns and the cross-scattering terms were removed to generate a correlation scattering factor (CSF). These CSFs were transformed through the generalized indirect Fourier transform (GIFT) method[33] to produce a distribution function, P(D$_{i,j}$), of the measured center-to-center distances between AuNPs. In 2013, Hura et al. were able to expand the application of this technique by probing the mismatch repair of MutS/L via XSI, demonstrating that the technique could be used to study DNA-protein interactions[34]. In this study 5nm AuNPs were prepared as previously described[35] and coupled to ssDNA via a Letsinger type trithiolated 5'-end to produce dsDNA-AuNPs conjugates with higher stability[36]. By using larger AuNPs the scattering signal from the gold dominated over cross-scattering terms by a factor of 1000, thus reducing the complexity of the analysis. In 2018, Zettl et al. were able to demonstrate the versatility of XSI by coupling 0.7 nm AuNPs to: (i) the ends of dsDNA ranging from 10- to 35-bp lengths, (ii) dsDNA molecules internally labeled, (iii) the ends of kinked-turn RNA constructs, and (iv) calmodulin, labeled at two lobes via cysteine residues[37]. This study then used *ab initio* mapping to determine the position of the gold labels and provide orientation in 3D space.

There are pros and cons to using different sizes of AuNPs as scattering points in XSI. Using AuNPs in the size range of 0.7 to 1.4 nm, as previously described,[29,30,37-39] produces interparticle distribution functions with narrow bandwidths, enabling distance measurements with higher

resolution, but since these AuNPs are on the size scale of the labeled macromolecules, the cross-scattering terms between gold and any other macromolecule must be mathematically removed, greatly complicating analysis, especially when studying DNA-protein dynamics. Alternatively, using larger, 5 nm AuNPs as previously described,[34] reduces the impact of cross-scattering terms and the complexity of the analysis, but also decreases the resolution and requires special care during the experimental design to ensure they do not perturb the conformational ensemble of the system being studied. Thus, for the proposed studies, a series of AuNPs ranging from 0.7 to 10 nm will be produced as previously described,[26,27] with smaller AuNPs being used for static structural studies with few interacting components and larger AuNPs used for dynamics studies. These AuNPs will then be coupled to experiment-specific ssDNA via a Letsinger type trithiolated 5'-end to produce single-stranded gold-DNA conjugates (Au-ssDNA) and annealed to complimentary strands to produce the desired conjugated dsDNA-AuNPs substrates as previously described.[34-36] In addition to creating dsDNA-AuNPs substrates, the smallest AuNPs can be conjugated to available cysteine residues, as previously described.[37,40,41] These smaller AuNPs would act as additional reference scatters and, after surmounting the inherent difficulties in analysis, would provide orientation information to aid in 3the D reconstruction of the system. In the work defined in this dissertation, we have created a modular toolbox of components in which to construct specially tuned dsDNA- or ssDNA-AuNP substrates to explore DNA-protein interactions and map the morphology of DNA-based bionanotechnologies.

1.4 Automated data collection, processing, and analysis pipeline

SEC-SAXS-MALS, HT-SAXS, and HT-XSI have all benefitted from the automated data collection, processing, and analysis pipeline made available through the interactive SIBYLS SAXS Processing (SSP) GUI,[12,42] designed to automate batch processing for SAXS and XSI data and provide real-time feedback. For SEC-SAXS-MALS, once samples are prepared, plated, and placed in the autosampler, collection can proceed as an individual sample, a single-user sequence, or a multiple-user sequence. Collection, processing, and deposition for SEC-SAXS-MALS at the SIBYLS beamline can all be managed through beamline-specific controls of the SSP GUI.

For HT-SAXS and HT-XSI, samples are loaded into a standard 96-well plate. Each sample is then transferred to the X-ray beam position via a Tecan Evo liquid handling robot with modified pipetting needles acting as sample cells and exposed as described previously.[43] Immediately upon finishing the first buffer bracketed sample region, the initial data processing begins by: i) circularly integrating the detector images to produce one-dimensional SAXS profiles, ii) normalizing the generated SAXS profiles to the diode current at the beamstop to compensate for minor fluctuations in the x-ray beam, and iii) subtracting the averaged scattering pattern of both bracketing buffers independently from each normalized sample curve. The final output of this process generates three sets of normalized, buffer subtracted SAXS curves for each sample. Two sets correspond to each bracketing buffer to ensure the subtraction process was not subject to instrument variations and a third which is an average between the original two sets to improve signal-to-noise. The data set is then reduced by sequentially averaging the scattering profiles over the 10 second exposure to produce high quality curves while eliminating the effects of radiation damage. For quality control, the variance between averaged SAXS profiles from the

two buffers is compared to identify three sources of error: (i) small bubbles in the needle sample cell, (ii) non-equivalent positioning of the needle in the x-ray beam, and (iii) temperature variation. Next, the individual frames from each set, are systematically checked for signs of radiation damage by calculating the variance between sequential frames over three ranges: (i) the Guinier region (low q ~0.01 to 0.05 Å^{-1}), (ii) the Porod region (q ~0.05 to 0.2 Å^{-1}), and (iii) the wide angle region (high q ~0.2 to 0.4 Å^{-1}). Region specific, frame averaged SAXS profiles are then produced for each set based on the values calculated for the number of radiation damage free frames and the ranges of each sliding region. The best scattering curves are selected and pairwise distribution functions, *P(r)*, are generated in batch using the automated GNOM[44] feature of the SSP GUI. Collection is run through Blu-Ice, but all other data processing is accomplished using our beamline-specific data processing pipeline by the SSP GUI. Further details pertaining to the development and application of the SSP GUI, and automated data processing pipeline are described throughout this dissertation.

1.5 Scope of Dissertation

This dissertation contains seven chapters, in addition to this introduction, concluding remarks, and an appendix of extra dissertation metadata at the end. Each chapter demonstrates the development and/or applications of advanced X-ray scattering techniques toward the exploration of DNA-protein interactions and engineering bionanotechnology. The following chapter summaries serve to provide relationships and transitions between the dissertation sections:

Chapter 2 - Size Exclusion Coupled Small Angle X-ray Scattering with Tandem Multi-Angle Light Scattering at the SIBYLS Beamline

This chapter serves to describe the design and development of the **S**tructurally **I**ntegrated **B**iolog**Y** for **L**ife **S**ciences (SIBYLS) beamline at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory in Berkeley, California, over the last five years. The chapter focuses on the development of size exclusion chromatography coupled small-angle X-ray scattering with in-line multi-angle light scattering (SEC-SAXS-MALS), a high-throughput, multimodal approach to studying structural biology with integrated *in situ* purification. This chapter also introduces the deployment of the SIBYLS SAXS Processing (SSP) GUI, an interactive software designed to automate batch processing for SAXS and X-ray Scattering Interferometry (XSI) data and provide real-time feedback.

Chapter 3 - Monitoring Nuclease Activity by X-ray Scattering Interferometry using Gold Nanoparticle Conjugated DNA

This chapter introduces X-ray scattering interferometry (XSI) and its development as a means of studying DNA damage repair and DNA-Protein interactions. The chapter uses the example of meiotic recombination 11 homolog 1 (MRE11) of which its nuclease activity is critical for multiple double-strand breaks (DSB) repair pathways as well as in replication. This chapter describes the methodology for monitoring nuclease activity and provides insight into DNA-protein interactions, demonstrating how the technique can aid in the development of inhibitors

that trap enzymes on the DNA substrate. Additionally, the chapter expands on the use of the SIBYLS SAXS Processing (SSP) GUI introduced in Chapters 2 and 3.

Chapter 4 - Defining the Molecular Basis of PARP-1 Damage Recognition, Activation, and Inhibition During Single-Stranded DNA Break Repair Through X-ray Scattering Interferometry

This chapter expands on the DNA damage repair capabilities of X-ray scattering interferometry (XSI) introduced in Chapter 3 using the example of the clinically important poly(ADP-ribose) polymerase 1 (PARP-1). The chapter investigates the damage recognition, activation, and inhibition of PARP-1 as it pertains to single-strand break (SSB) repair. This chapter also describes overlapping designs involving a merger of XSI and the size exclusion chromatography coupled with small-angle X-ray scattering with in-line multi-angle light scattering (SEC-SAXS-MALS) techniques described in Chapter 2.

Chapter 5 - Quantitative Protein Corona Composition, Driving Forces, and Dynamics on Carbon Nanotubes in Biological Environments

This chapter expands the breadth of the dissertation into the realm of bionanotechnology engineering and provides context for the work in Chapter 6. In this chapter we demonstrate how conventional small-angle X-ray scattering (SAXS) can be used to explore protein corona formation on DNA-functionalized single-walled carbon nanotubes (ssDNA-SWCNTs), a nanotechnology used for nanosensing and delivery, in blood plasma and cerebrospinal fluid. This chapter underlines the inadequacies associated with traditional SAXS in studying these systems and paves the way for the further development of the X-ray scattering interferometry (XSI) techniques seen in Chapter 6.

Chapter 6 - Mapping the Morphology of DNA on Carbon Nanotube-Based Sensors in Solution using X-ray Scattering Interferometry

This chapter serves as the zenith of this dissertation in demonstrating a novel and powerful use of X-ray scattering interferometry (XSI) in providing insights into the nanosensing mechanisms and information for future design strategies for DNA-functionalized SWCNT technologies like those introduced in Chapter 5. The chapter demonstrates direct in-solution measurement of ssDNA geometries on SWCNTs via XSI previously unobservable but conventional means.

Chapter 7 - Additional Cases for Size Exclusion Coupled Small Angle X-ray Scattering

This chapter serves to demonstrate four case studies for the size exclusion chromatography coupled SAXS with in-line multi-angle light scattering (SEC-SAXS-MALS) system introduced in Chapter 2. These cases were selected to best highlight the robust capabilities of this technique bringing together concepts from previous chapters.

2 Size Exclusion Coupled Small Angle X-ray Scattering with Tandem Multi-Angle Light Scattering at the SIBYLS Beamline[†]

2.1 Chapter Abstract

Small-angle X-ray Scattering (SAXS) has been a versatile technique for studying biomolecules in solution for several decades now. Developments in SAXS techniques that integrate *in situ* purification with a high-throughput, multimodal design philosophy have revolutionized the reach and tempo of BioSAXS experiments. The current zenith of the field comes in the form of size exclusion chromatography coupled SAXS with in-line multi-angle light scattering (SEC-SAXS-MALS). This technique has been a considerable focus at the **S**tructurally **I**ntegrated **B**iolog**Y** for **L**ife **S**ciences (SIBYLS) beamline at the Advanced Light Source (ALS) in Berkeley, California, over the last five years and continues to be a point of active development. In this chapter, we describe the design of the SEC-SAXS-MALS system and general guidelines for collecting, processing, and analyzing SEC-SAXS-MALS data at the SIBYLS beamline.

2.2 Introduction

Small-angle X-ray Scattering (SAXS) is a convenient and versatile technique for studying the structure, complexation, and activity of biomolecules in solution [1]. The method utilizes X-ray radiation to observe changes in the electron density of macromolecules as they exist in solution at near biological conditions. While commercial, "tabletop" SAXS instruments are becoming more widely available, the most popular method still involves the use of synchrotron facilities like the Advanced Light Source (ALS) in Berkeley, California, due to the significantly increased flux of the X-ray source. This increased brightness consequently reduces the exposure times needed from minutes or hours down to seconds. Leveraging the critical reduction in data collection time, SAXS beamlines at synchrotrons worldwide have been able to expand upon traditional SAXS collection methods to develop new techniques with a high-throughput, multimodal design philosophy.

In 2001, the **S**tructurally **I**ntegrated **B**iolog**Y** for **L**ife **S**ciences (SIBYLS) beamline (bl12.3.1)[2] was commissioned at the ALS to be the flagship BioSAXS[1] beamline at Lawrence Berkeley National Laboratory (LBNL). This beamline specializes in technologies for the characterization of biological molecules and has been a staging ground for the development of integrated biophysical techniques including size exclusion chromatography coupled SAXS with in-line multi-angle light scattering (SEC-SAXS-MALS) [3-8] and high-throughput SAXS (HT-SAXS) [9,10]. Both systems are capable of automated data collection, with SEC-SAXS-MALS having the added benefits of providing *in situ* purification along with in-line complementary spectroscopic and light scattering techniques. These techniques provide immediate validation for the analogous values they provide. The molecular weight calculated from the MALS (MW$_{MALS}$) is analogous to the molecular weight calculated from the SAXS intensity at the zero angles, I(0), (MW$_{SAXS}$). One

---

[†] Published as Rosenberg, D. J.; Hura, G. L.; Hammel, M. Size Exclusion Chromatography Coupled Small Angle X-Ray Scattering with Tandem Multiangle Light Scattering at the SIBYLS Beamline. In *Methods in Enzymology*; Elsevier, 2022; Vol. 677, pp 191–219.

angle from the MALS is isolated to generate quasi-elastic light scattering (QELS, a.k.a. dynamic light scattering) data providing values for the radius of hydration (Rh) analogous to the radius of gyration (Rg) calculated from the SAXS. The continued development of SEC-SAXS-MALS builds upon an expanding pool of previous work [13-21]. In this chapter, we first describe the design of the current SEC-SAXS-MALS system at the SIBYLS beamline in detail. Then we discuss the collection and processing of SEC-SAXS-MALS data specific to the SIBYLS beamline and outline general protocols for the analysis of SEC-SAXS-MALS data.

2.3   Endstation Design

The design of each beamline and endstation at an accelerator facility is uniquely designed and often meant to serve a very specific function, striving to enable experiments that often cannot be accomplished in any other way [45-47]. The SIBYLS beamline is no exception to this general design philosophy striving principally to achieve high collection speeds and robust automatization to study structural biology via SAXS. For the SIBYLS beamline to achieve automated high-throughput data collection, the conventional SAXS endstation had to be modified to include several significant design changes (Figure 2-1). Firstly, a modern shutterless detector (Dectris PilatusX3 2M, Baden, Switzerland) was required to increase framerates and expand the detector area. Additionally, to incorporate the changes to the larger detector area and increase the sample-to-detector distance for higher resolution at lower scattering angles, a new vacuum flight tube was designed. This flight tube includes a 2 x 2 cm X-ray transparent 20 µm potassium aluminosilicate (Muscovite Mica) window (Goodfellow Corp., Coraopolis, PA) pressure sealed with rubber O-rings to cover the 1 cm diameter entrance to the flight tube and a sheet of 127 µm polyimide (Kapton) film (McMaster-Carr, Santa Fe Springs, CA) pressure sealed with 318 mm thick oil-resistant nitrile (Buna-N) rubber (McMaster-Carr, Santa Fe Springs, CA) to cover the 29.5 x 16.6 cm exit. In order to slow the inevitable deflection of the Kapton it would experience being subjected to ~4900 N and to reduce the air gap between the exit of the flight tube and the detector, horizontal bars are built into the flight tube which is identical in size and aligns to the inter-module spaces on the detector face (Figure 2-1 bottom inset).

The X-ray energy is typically set at 10keV (λ=1.24 Å) but can be changed between ~7-16 keV with 110 (E/ΔE) resolution using a monochromator with dual Mo/B$_4$C multilayers as previously described [2]. The standard working energy of 10keV and an increased sample-to-detector distance of 2075 mm, the resulting scattering vector (q) ranges from 0.008 Å^{-1} to 0.46 Å^{-1}, which corresponds to the 800Å to 14Å resolution in real space. The scattering vector is defined as $q = 4\pi sin\theta/\lambda$, where 2θ is the scattering angle. Secondly, the integration of a modified liquid handling robot with custom-designed liquid handling needles was leveraged to establish the current HT-SAXS system [9,10]. Finally, a system designed from carefully selected commercial instruments (see Section 2.3.1) was adapted in conjunction with a purpose-built sample flow cell (see Section 2.3.2) to establish the current SEC-SAXS-MALS system (Figure 2-1).
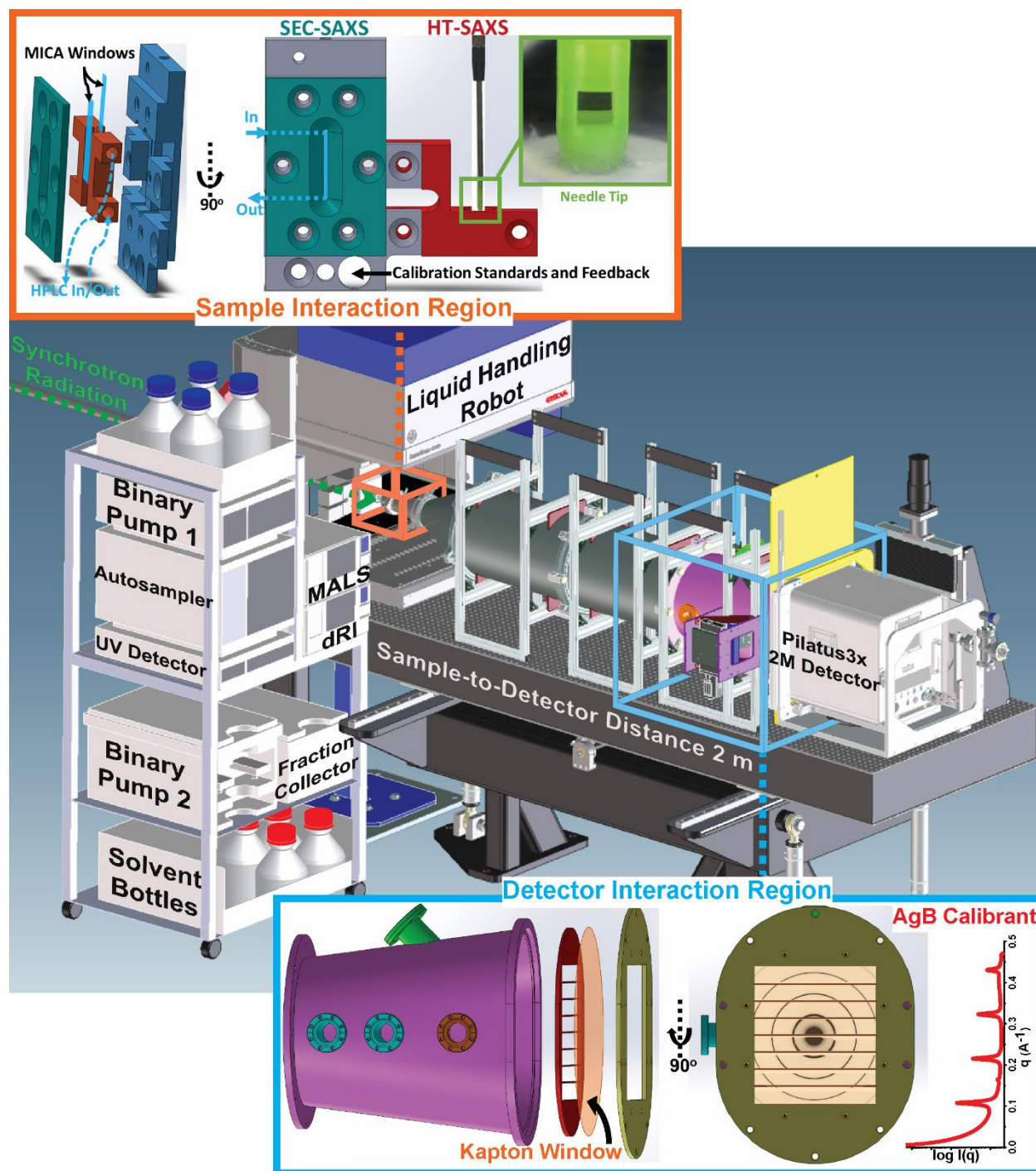
Figure 2-1. Beamline layout showing a schematic representation of the SIBYLS beamline with SEC-SAXS-MALS and HT-SAXS systems. Top inset (orange) shows the sample interaction region with SEC-SAXS flow cell and HT-SAXS needle designs. Bottom inset (blue) shows the detector interaction region with flight tube design and representative silver behenate calibration showing scattering vector.

2.3.1   SEC-SAXS-MALS Layout

The SEC-SAXS-MALS system has both a primary flow path for the active experiment and a secondary flow path to simultaneously equilibrate the system for the next running conditions (Figure 2-2). Both flow paths are modular but under standard running configuration, both flow paths consist of four available buffer slots. Buffer selection and flow rate of each flow path is controlled by binary pumps from an Infinity series high-performance liquid chromatography (HPLC) system (Agilent, Santa Clara, CA). In the primary flow path, a sample is injected from a standard 96-well plate using a dual-needle, 1290 Infinity series thermostat-controlled autosampler (Agilent, Santa Clara, CA) set to 7 °C. The elution is separated by a selected size-exclusion column and the absorbance is monitored by an in-line 1290 Infinity II Diode Array Detector (DAD) (Agilent, Santa Clara, CA). Subsequently, in-line MALS and QELS experiments are performed using an 18-angle DAWN HELEOS II (Wyatt Technology, Goleta, CA) light scattering detector with integrated QELS detector. The sample then travels to the custom flow cell (see Section 2.3.2, Figure 2-1, and Video 1 (only in published version)) for SAXS measurements then back to an Optilab T-rEX (Wyatt Technology, Goleta, CA) refractometer to measure the differential refractive index (dRI) (Wyatt Technology, Goleta, CA) needed for proper MALS analysis. Finally, the sample is either sent to waste or recollected in a 1260 series thermostat-controlled fraction collector (Agilent, Santa Clara, CA) set to 4 °C. As the primary flow path experiment proceeds, the secondary flow path has been equilibrating another
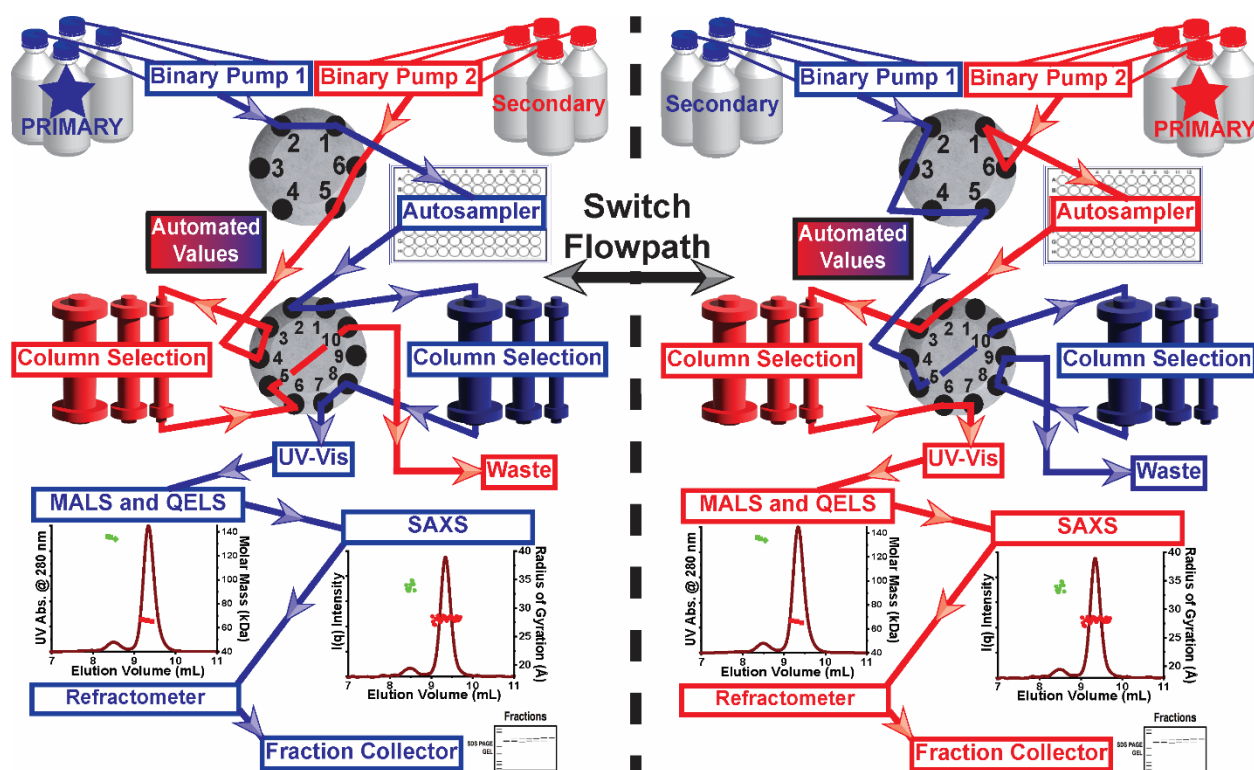
Figure 2-2. Detailed SEC-SAXS-MALS layout demonstrating the two switchable flow paths made possible by the automated valve assembly.

column for the subsequent experimental conditions and upon completion of the active experiment, switches over to being the primary flow path through the control of a custom dual-valve system (Agilent, Santa Clara, CA) outlined in Figure 2-2.

2.3.2 SEC-SAXS Flow Cell Design

When collecting SEC-SAXS-MALS data under standard operating conditions, a custom, modular flow cell is used (Figure 2-1 top inset and Video 1 (only in published version)). This flow cell utilizes X-ray transparent 20 μm potassium aluminosilicate (Muscovite Mica) windows (Goodfellow Corp., Coraopolis, PA) held on either side of a 1.2 x 1.2 x 20 mm channel by pressure and sealed with 100 μm polytetrafluoroethylene (PTFE, Teflon) gaskets. The flow cell is under continued development and the current version is designed to operate as a standard flow cell with a single inlet and outlet with a calculated volume of 28.8 μL.
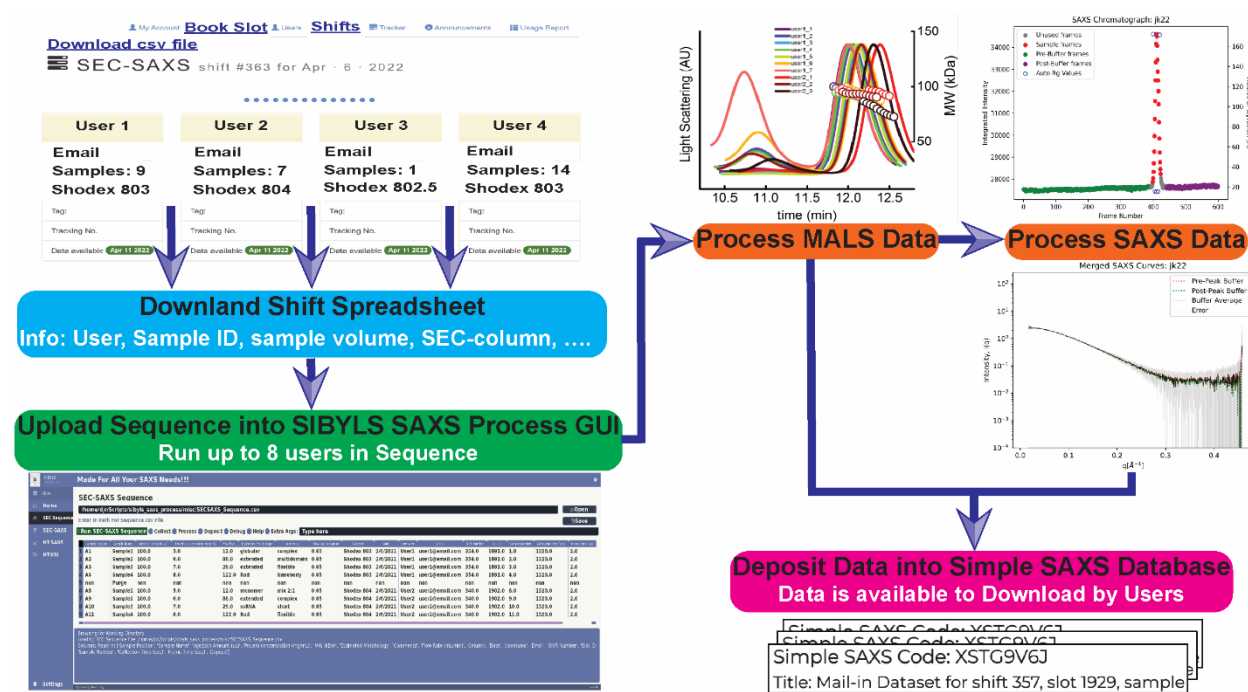
Figure 2-3. Representation of the automated multi-user sequence data collection, processing, and deposition pipeline.

2.4 Sample Preparation for SEC-SAXS-MALS

The overall pipeline for the mail-in SEC-SAXS-MALS program, data collection, processing, and deposition is shown in Figure 2-3. This process begins with users booking a slot on our website and preparing samples to be received and collected by our beamline staff. A major advantage of the SEC-SAXS-MALS over other traditional SAXS techniques is the *in situ* purification, accomplished through SEC, however this comes at the cost of sample dilution. Thus, the addition

of a pre-purification step, specific to the macromolecule of interest, helps preserve the integrity of the system and allows for a more accurate initial concentration measurement (see Note 1). While generating and purifying macromolecules is case-specific, some general recommendations on sample purity and concentration in addition to selecting the elution buffer and column conditions should be followed when preparing for a SEC-SAXS-MALS experiment [4,6,8,48-52].

2.4.1   Equipment

- Pipette
- Pipette tips
- 96-well plate (see Note 2)
- 96-well plate adhesive pierceable polyethylene films
- Eppendorf tubes
- Parafilm®
- 50 mL falcon tubes
- Temperature controlled centrifuge compatible with Eppendorf tubes (see Note 3)
- 500 mL Agilent HPLC compatible solvent bottles
- Solvent bottle filtration device
- Solvent bottle 0.22 μm membrane filters

2.4.2   Reagents

- A sample with a minimum volume of 50 μl (see Note 4) at proper concentration (see Note 1).
- A minimum of 50 ml of 10x SEC buffer (see Note 5)
- Deionized water (diH$_2$O)

2.4.3   Procedure

1. To collect SEC-SAXS-MALS data at the SIBYLS beamline, a user will first book a time slot on the SIBYLS mail-in website (see Note 6) using the "Book Slot" link on the top of the page (Figure 2-3). The number of samples and buffer conditions are subject to limitations and availability (see Note 7).

2. Macromolecules are produced and purified by sample-specific methods.

3. The appropriate sample concentrations and volumes are pipetted into individual Eppendorf tubes and sealed with Parafilm® and placed in 50 mL falcon tubes for shipment.

4. A minimum of 50 ml of 10x of desired SEC buffer (see Note 5) is prepared in 50 mL falcon tubes and sealed with Parafilm® for shipment.

5. Sample, buffer, and column information (see Note 8) is added to booked slot.

6. Samples and buffer are packed for shipping to the ALS (see Note 9).

7. Once samples are received, the 50 ml of 10x of desired SEC buffer is added to 450 mL diH$_2$O to prepare the 1x buffer used as the SEC mobile phase.

8. 1x SEC buffer is filtered with 0.22 µm membrane filters into a 500 mL Agilent HPLC compatible solvent bottle.

9. Received samples are thawed rapidly at room temperature if frozen and then immediately placed in a temperature-controlled centrifuged at 4 °C (unless samples require a different storage temperature).

10. Samples are spun at 8000 rpm for 5 mins at 4 °C to remove macromolecule precipitation that might have occurred during transport.

11. Centrifuged samples are added sequentially to the 96-well plate in the order of the user sample information, sealed with 96-well polyethylene films, and placed in the Agilent autosampler at 7 °C (unless samples require a different storage temperature) before data collection.

2.4.4   Notes

1. The sample concentration should be 3-10 mg/ml with smaller molecules requiring higher concentrations. The general rule of thumb for concentration uses the following general formula for proteins and about half the concentration for nucleic acids:

$$concentration\ (mg/mL) = \frac{300}{molecular\ weight\ (kDa)} \qquad \text{2-1}$$

For example, a 100 kDa protein = 3 mg/mL minimum recommended concentration.

2. For a 96-well plate, we recommend Axygen® 96-well Polypropylene PCR Microplate, but others should be fine if they are listed as compatible with a 1290 series Agilent Autosampler.

3. If smaller Eppendorf tubes or PCR tubes are used, centrifuge adaptors may be required.

4. To avoid bubbles in the system and account for inconsistencies in pipetting the sample into the 96-well plates, 5 µL less is injected than is put into each well of the 96-well plate. The maximum injection volume is 100 µL limited by the autosampler loop size.

5. Special consideration should be taken when preparing the buffer for a SAXS experiment [48]. Ideally, the sample should be prepared in a buffer that is identical or very similar to the SEC buffer, but this is not required as the buffer will exchange on the column as the elution proceeds. For SEC buffers, salt concentrations ≤ 500 mM and pH ≤ 8 should be maintained when using the standard Shodex columns provided by the SIBYLS beamline as per the manufacture's recommendations. The concentration of glycerol should not exceed 2% and

the concentrations of other stabilizers or additives should be kept to a minimum to reduce the contribution to the background scattering and avoid high column pressure.

6. SIBYLS mail-in website and most current instructions can be found here: https://bl1231.als.lbl.gov/htsaxs/

7. Each beamtime slot is currently limited to 15 samples using 1 column/buffer condition, but these numbers are subject to change and can be modified upon request.

8. Please specify the columns to use in your experiments. If standard columns do not meet the needs of the experiment, other columns may be available or may need to be sent with samples. A list of available columns with a summary of their properties is shown in Table 2-1.

Product Name	Exclusion Limit (Protein kDa)	Particle Size (μm)	Maximum Pore Size (Å)	Column Size (mm) ID x L
PROTEIN KW-802.5	150	5	400	8 x 300
PROTEIN KW-803	700	5	400	8 x 300
PROTEIN KW-804	1000	7	1500	8 x 300

Table 2-1. Available size-exclusion columns.

9. We recommend that samples are shipped in standard Eppendorf PCR tubes together with 10x buffer at 4 °C with cold packs or frozen with dry ice in a Styrofoam shipping container. DO NOT ship wet ice. Clearly label the outside of your shipping box with the temperature of your samples. This is extremely important if a shipping delay occurs. Lastly, please provide us with a shipping tracking number. See SIBYLS mail-in website (see Note 6) for address.

2.5 Data Acquisition

Once samples are prepared, plated, and placed in the Agilent autosampler, SEC-SAXS-MALS collection can proceed as an individual sample, a single user sequence, or a multiple user sequence. Collection, processing, and deposition for SEC-SAXS-MALS at the SIBYLS beamline can all be managed through beamline specific controls software called SIBYLS SAXS Process following the pipeline shown in Figure 2-3.

2.5.1 Equipment

- Pilatus 3X 2M Detector (Dectris, Baden, Switzerland)

13

- 1290 Infinity Multisampler with Sample Loop-Flex 100 μL and InfinityLab Sample Thermostat (Agilent, Santa Clara, CA)

- 1290 Infinity II High Speed Pump (Agilent, Santa Clara, CA)

- 1290 Infinity II Diode Array Detector (DAD) (Agilent, Santa Clara, CA)

- 1260 Infinity Fraction Collector with Sample Thermostat (Agilent, Santa Clara, CA)

- 1260 Infinity Binary Pump with four solvent head (Agilent, Santa Clara, CA)

- In-line degasser (Agilent, Santa Clara, CA)

- 2pos/6port valve head (Agilent, Santa Clara, CA)

- 2pos/10port Quick Change valve head (Agilent, Santa Clara, CA)

- 18-angle DAWN HELEOS II multi-angle light scattering (MALS) detector with WyattQELS unit installed (Wyatt Technology, Goleta, CA)

- Optilab T-rEX differential Refractive Index (dRI) detector (Wyatt Technology, Goleta, CA)

- Trident LC Column in-line filter (Restek, Bellefonte, PA)

- Three standard Shodex KW series columns (Showa Denko, Tokyo, Japan)

- Agilent OpenLab (version Rev C.01.10 or later recommended) (Agilent, Santa Clara, CA)

- Wyatt ASTRA software (version 7.1 or later recommended) (Wyatt Technology, Goleta, CA)

2.5.2 Reagents

- A minimum of 500 ml of 1x SEC buffer

- Deionized water (diH$_2$O)

2.5.3 SEC-SAXS-MALS Sequence Setup

1. Once user samples are received, Agilent and Wyatt systems are turned on and the modules with lasers are allowed to "warm up" (see Note 1).

2. SEC is then performed using one of three standard Shodex KW series columns (Showa Denko, Tokyo, Japan) depending on the expected size range of the macromolecules in question (see Section 2.4.4 Note 8). This column is connected from a selection of columns found in the chilled column chamber (Figure 2-2) using standard practices (see Note 2).

3. The column is equilibrated ahead of time with a mobile phase (1x buffer) at a standard flow rate of 0.65 mL/min for about 1 hour until the differential refractive index (dRI) signal stabilizes.

4. On mail-in SAXS website (see Section 2.4.4 Note 6), navigate to the current shift using the "Shift" link on the top of the page (Figure 2-3) and download the spreadsheet with all user

and sample information as a comma-separated value file (.csv) using the "Download csv file" link in the upper right corner of the shift page (Figure 2-3).

5. Once the system is at equilibrium, use the downloaded shift spreadsheet to set up complimentary sample sequences in the Agilent OpenLab software, the Wyatt ASTRA software (see Note 3), and the SIBYLS SAXS Process GUI (see Section 2.5.4).

6. The sample is separated by SEC and the elution is monitored with up to eight discreet wavelengths (see Note 4) by an in-line DAD to monitor sample concentration as it elutes off the column.

7. The sample then travels through the in-line MALS to obtain MW_{MALS} and Rh measurements and onto the SAXS sample interaction region.

8. After the sample passes through the x-ray beam, it travels to the in-line refractometer to measure reflectivity index (dRI) required to calculate the MW_{MALS} from the MALS data (see Note 5).

9. Finally, the sample will either be sent to waste or recollected in the fraction collector as the experiment demands (see Note 6).

10. A typical sample will require approximately 25 min to complete data acquisition and, if needed, an additional 35 min system purge to establish new equilibrium conditions for each unique user or buffer condition.
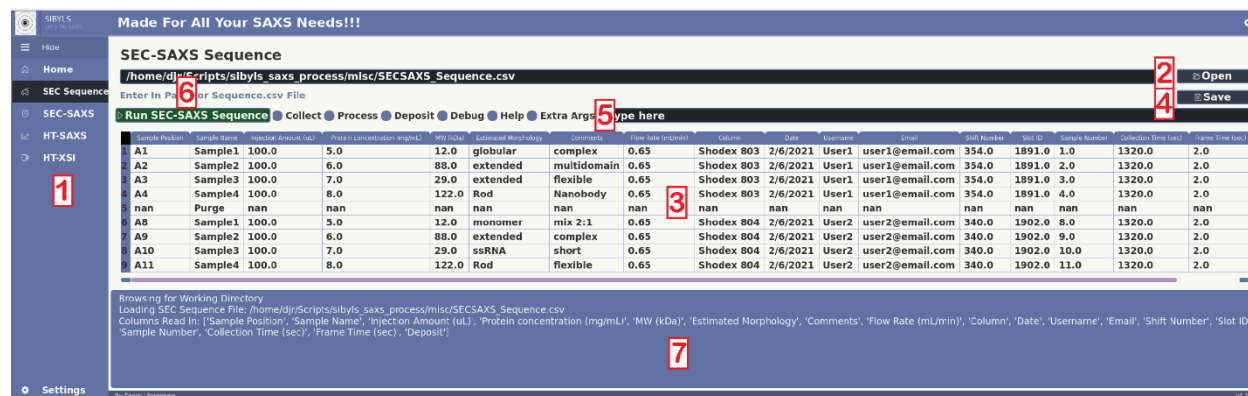
Figure 2-4. SIBYLS SAXS Process GUI showing the SEC-SAXS automated multi-sample sequence data collection, processing, and deposition panel.

2.5.4   Setting Up the SIBYLS SAXS Process GUI

1. Once the system is at equilibrium and the complimentary sample sequences are set up in the Agilent OpenLab software, the Wyatt ASTRA software (see Note 3), the SIBYLS SAXS Process GUI (see Note 7) is used to set up the automated multi-sample sequence data collection, processing, and deposition.

2. After loading the SIBYLS SAXS Process GUI, the main side menu is expanded (Figure 2-4 #1) and the "SEC Sequence" tab is selected to open the multi-sample sequence widget.

3. The "Open" button (Figure 2-4 #2) is used to load a pre-made sample sequence spreadsheet (see Note 8) into the editable sequence table widget (Figure 2-4 #3). For details on a typical sequence setup, see Note 9.

4. Once loaded, the sequence can be edited in the table widget (Figure 2-4 #3) and saved using the save button (Figure 2-4 #4). If changes are made to the sequence, they must be saved before continuing (see Note 10).

5. If multiple users with different mobile phases or column conditions are in a single sequence, "Purge" is placed in the appropriate row in the "Sample Name" column, so the program knows to purge and equilibrate the column between samples (see Note 11).

6. The option buttons (Figure 2-4 #5) select for specific actions or arguments during the sequence run (see Note 12). For example, the user can select "Collect", "Process", or "Deposit" data using the respective option buttons.

7. To accurately automate SEC-SAXS data processing for multiple users, a mask file and an experimental parameters file must first be created (see Note 13).

8. Once the sequence is saved and the options are selected, the sequence can be initiated by hitting "Run SEC-SAXS Sequence" (Figure 2-4 #6). Do this only when the complementary sequences are prepared and started on the in-line Agilent and Wyatt systems (see Note 14). The run can be followed in the output terminal (Figure 2-4 #7).

9. Once the sequence has completed and all data has been collection and processed, the results are organized in the following file hierarchy for each Shift Date -> Username -> Sample:

Date_Folder (i.e., /data/secsaxs/041622):
```
└─ Results
   └─ 2022_04_16_SEC-SAXS_Username1
      └─ Sample1
         ├─ Sample1_Peak_Finding (Auto Peak Finder Results)
         │  ├─ Sample1_LC_Trace.png (SEC-SAXS Chromatograph)
         │  ├─ Sample1_Merged_Curves.png (Merged SAXS Curves)
         │  ├─ Subtracted (Subtracted SAXS Curves for Each Frame)
         │  │  ├─ Full_Buffer
         │  │  ├─ Post_Buffer
         │  │  └─ Pre_Buffer
         │  └─ Unsubtracted (Unsubtracted SAXS Curves)
         │     ├─ Buffer_Frames
         │     └─ Sample_Frames
         └─ Sample1_Results (All Unsubtracted SAXS Curves)
```

2.5.5   Notes

1. The Agilent Diode Array Detector (DAD) and the Wyatt MALS detector should take approximately 30 min and 60 min respectively to "warm up" as recommended by the manufacturer's instructions. The MALS is ready when the 90-degree angle fluctuates < +/- 0.002 V.

2. Initialize Agilent pumps prior to connecting desired column in order to bleed air out of the lines and establish a wet connection.

3. See user manuals for details on setting up sequences on Agilent OpenLab (version Rev C.01.10 or later recommended) controlling the Agilent HPLC modules and Wyatt ASTRA software (version 7.1 or later recommended) controlling the Wyatt instruments.

4. In a typical structural biology experiment, proteins and RNA are monitored at 280 nm, and DNA is monitored at 260 nm, however some macromolecules may require additional or alternative wavelengths to be monitored.

5. The dRI is necessary to determine fluctuations in the mobile phase, which may lead to differences in the motion of the particles measured by MALS and is thus needed to make corrections in the MW_{MALS} calculations.

6. Fraction collection is not a standard feature of a typical SEC-SAXS-MALS experiment and is usually only enabled upon request.

7. The SIBYLS SAXS Process package is available on our GitLab repository (https://git.bl1231.als.lbl.gov/saxs/sibyls_saxs_process).

8. Pre-made sample sequence spreadsheets can be loaded or saved either a comma-separated value file (.csv) or an excel workbook file (.xlsx).

9. A standard template should be followed for a typical sequence, which comes with the SIBYLS SAXS Process package found on our git repository (see Note 7) following the path: /sibyls_saxs_process/misc/SECSAXS_Sequence.csv

10. If changes are made to the sequence, the program will not allow you to start the sequence until all changes have been saved. The terminal will prompt you if there is a conflict (Figure 2-4 #7).

11. A minimum system purge time of 35 min is typically required for the system to reach equilibrium when a change in mobile phase or column is required. These "purge" breaks need to be set up independently in the Agilent OpenLab software, the Wyatt ASTRA software (see Note 3), and the SIBYLS SEC-SAXS Process GUI.

12. To add additional, non-standard arguments the "Extra Args" options button can be selected, and extra inputs typed into the text box next to it. Use the help option for more details on other arguments that can be used.

13. To automate data processing, the "Process" option button must be selected (Figure 2-4 #5). Additionally, a mask for the unused regions of the detector (mask.ASC) and an experimental parameters file (ExptParam) containing information on the X-ray wavelength (Å) and the detector metrology must be created. These files are generated using the FIT2D software[53]. These two files must be placed in the date directory for the shift where all current user data is being collected (i.e., /data/{*account collecting data*}/{*shift date*}). If no files are found at that location the program defaults to pulling the necessary files from the last ones generated but this might lead to processing error and the data may need to be re-processed.

14. The Wyatt ASTRA, the Agilent OpenLab (see Note 3), and the SIBYLS SAXS process sequences should be started in this order respectively.

2.6   Data Processing

In a typical SEC-SAXS-MALS experiment, the depth and detail of the analysis may vary but the most basic processing and analysis of the raw data typically remains the same. Indirectly the complementary spectroscopy and light scattering measurements are used to monitor the concentration and homogeneity of the samples in real time to ensure the quality of the sample immediately before SEC-SAXS measurements are taken. More directly however, some rudimentary tools are included in the SIBYLS SAXS Process GUI to automate SEC-SAXS data processing and display the most critical information to provide rapid feedback on data quality. Certain standard practices should be taken to ensure the quality and accuracy of SEC-SAXS-MALS data and both sources of feedback end up being crucial towards meeting this goal.

2.6.1   Processing MALS, QELS, and UV Data

1. System normalization and calibration are performed for each new experimental condition with bovine serum albumin (BSA) (see Note 1) or some other protein of known MW.

2. The UV-vis signal from the DAD is used to calculate the concentration of each eluted sample by setting up a standard curve in either Agilent OpenLab or Wyatt ASTRA software (see Section 2.5.5 Note 3).

3. The MALS and dRI signals are used to determine MW_{MALS} and the QELS measurements are used to determine the Rh across the principal peaks in the SEC analysis all processed and analyzed with the Wyatt ASTRA software (see Section 2.5.5 Note 3).

4. All spectroscopy and light scattering measurements can be used to monitor the concentration and homogeneity of the sample across the elution peaks in real-time and serve as an early prediction of the quality of the SEC-SAXS data being collected.

2.6.2   Processing SEC-SAXS Data

1. Beamline-specific software built into the SIBYLS SAXS Process GUI is used to process raw SAXS-detector images for individual samples by expanding the main side menu (Figure 2-5 #1) and selecting the "SEC-SAXS" tab to open the SEC-SAXS processing widget. Alternatively, SEC-SAXS data processing can be run in batch or automated to proceed each sample as data is collected using the "SEC-SAXS sequence" tab and selecting the "Process" option as described (see Section 2.5.4 and 2.5.5 Note 13).

2. To process data individually within the "SEC-SAXS" tab, the "Open" button (Figure 2-5 #2) is used to load the directory containing the raw SAXS detector images (.cbf file format).

3. To accurately process SAXS data, a mask file and an experimental parameters file must first be created (see Section 2.5.5 Note 13).

4. After selecting the working directory (Figure 2-5 #2), clicking the "Process SEC-SAXS" button (Figure 2-5 #3) with the desired optional argument buttons (Figure 2-5 #4) (see Section 2.5.5 Note 12) will produce SAXS curves for every exposure frame by radially integrating and normalizing each raw detector image, as previously described [9,10]. The output of this process is a

results folder with the name *samplename_*Results, within the loaded working directory (Figure 2-5 #2) (see Note 2).

5. The data files produced by the "Process SEC-SAXS" step each represent a single time point or frame, usually 1-3 sec long, during the SEC elution. The "Auto Peak Finder" button (Figure 2-5 #5) sequentially integrates the SAXS curve for each frame across the elution and display the results as a SEC-SAXS chromatogram (Figure 2-5 #8). The automatic peak finder function by default then takes each frame selected as the primary sample elution peak, subtracts an averaged selection of both pre- and post-peak buffers from each frame, and merges the results leading to three final curve (see Note 3). Alternatively, to the default automated peak finding, users can customize the peak finding parameters by modifying the number of peaks found, sample frames merged, and buffer frames averaged by using the associated optional arguments (Figure 2-5 #4) (see Section 2.5.5 Note 12).
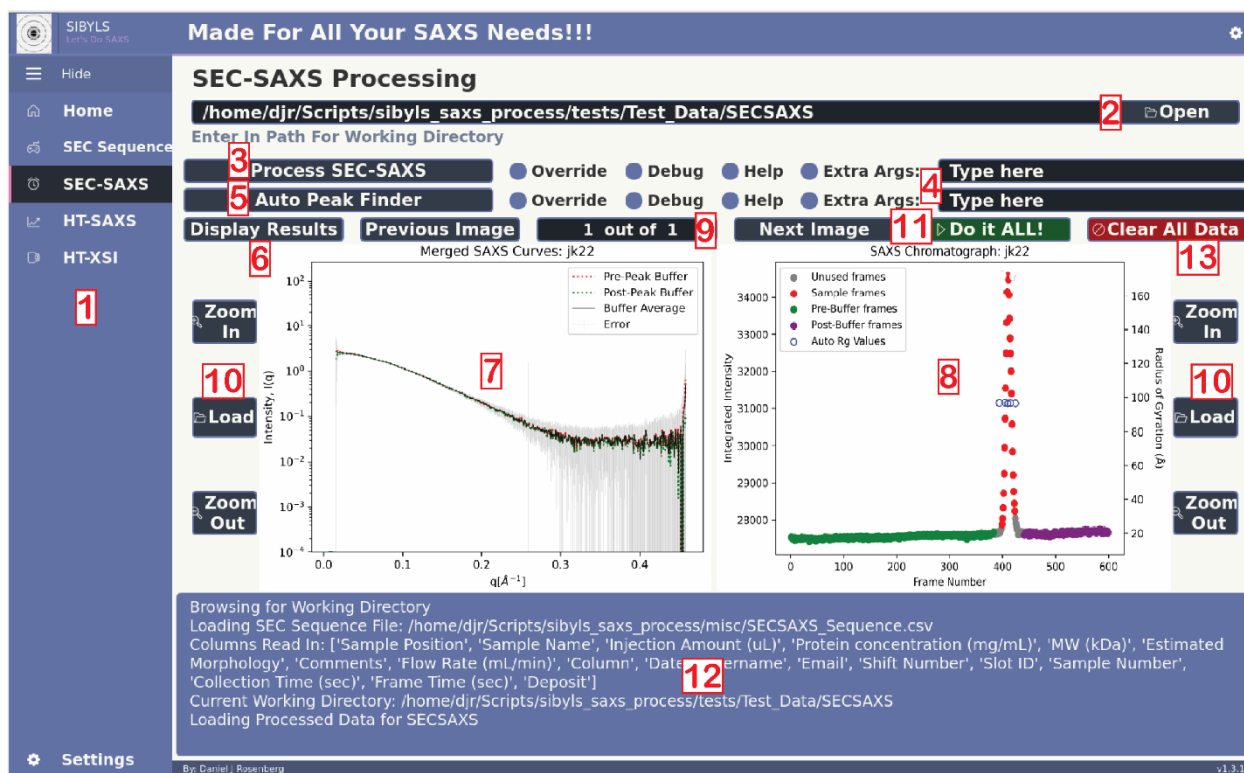
Figure 2-5. SIBYLS SAXS Process GUI showing the SEC-SAXS single sample data processing panel with automated peak finding.

6. The "Display Newest Results" button (Figure 2-5 #6) shows the latest results from the "Auto Peak Finder" function as a set of two corresponding plots showing the three final subtracted SAXS curves (see Note 3) (Figure 2-5 #7) and the SEC-SAXS chromatogram (Figure 2-5 #8) that produced them. All sets in the same directory can be browsed with the "Previous" and

"Next" image buttons (Figure 2-5 #9) and older sets can be searched and displayed with the "Load" buttons (Figure 2-5 #10).

7. To automate the whole SEC-SAXS data processing pipeline for a single sample, use the "Do it ALL!" button (Figure 2-5 #11). The output of each step is displayed in the terminal below (Figure 2-5 #12).

8. Any processed data in the working directory (Figure 2-5 #2) can also be removed by using the "Clear All Data" button (Figure 2-5 #13) and answering 'yes' to confirm when prompted in the pop-up window (see Note 4).

2.6.3 Notes

1. BSA is typically collected by injecting 50 µL of 7 mg/mL sample in the same mobile phase and on the same column as the experimental conditions demand. The use of a protein of known MW as a standard is particularly critical for normalizing the MALS system for the new mobile phase and column but is also useful to ensure the quality of the acquired SEC-SAXS data before starting to run important samples.
2. The radially integrated and normalized data files (.dat) are in the form of three-column, space separated files. Column one is the q values (Å^{-1}), column two the Intensity values I(q), and column three the error in the I(q) [1].
3. The three final curves produced by the default automatic peak finder are: 1) pre-peak buffer subtracted, 2) post-peak buffer subtracted, and 3) an average of the two previous curves.
4. The data removed with the "Clear All Data" button (Figure 2-5 #13) is not permanently removed and are instead moved to a temporary "Trash Can" folder within the working directory (Figure 2-5 #2) in case the user wished to reverse the removal. This is only done once, however, and the subsequent removal of data by the same name will override it.

2.7   Data Validation and Analysis

The steps shown previously in Section 2.6 are meant for preliminary data processing and quickly displaying results to monitor data quality while running the SEC-SAXS-MALS system. While the SIBYLS SAXS Process interface is currently designed for running the system and feedback during the experiments, continued development is underway to expand the automated data processing pipeline. There are several computational tools available now to further analyze SEC-SAXS data including SCÅTTER [54], RAW [55], and CHROMIXS [56]. These software options are not discussed here, and the focus is instead on the theory of how to extract, validate, and analyze the information obtained from a typical SEC-SAXS-MALS experiment.

Whether you are using SIBYLS SAXS Process or any other software for analyzing SEC-SAXS-MALS data, certain factors must be kept in mind when assessing the quality and validity of your measurements. For SEC-SAXS-MALS data, this validation generally starts with the assessment of the SEC-SAXS chromatogram, and then by analyzing the final buffer subtracted, merged SAXS curves. Subsequently, the SEC-SAXS-MALS data is further validated by comparing analogous values from the complementary spectroscopic and light scattering techniques collected in-tandem with the SAXS. In combination, the in-line UV, MALS, and QELS measurements

provide an accurate assessment of sample aggregation, heterogeneity, and the oligomerization state of macromolecules across the SEC elution and can often aid in SEC-SAXS data interpretation.

2.7.1   Self-validation and Analysis of SEC-SAXS Data

For SEC-SAXS data, validation generally starts by selecting baseline frames corresponding to the buffer to use as subtraction for the sample peaks. The baseline of a SEC-SAXS chromatogram can vary significantly depending on many factors including fluctuations in the X-ray beam intensity or position, sample concentration, and the sensitivity of the sample to radiation. For a well-behaved sample, run on a fully equilibrated system (Figure 2-6A), the choice of buffer subtraction is relatively straightforward a small, stable buffer region (See Note 1), typically immediately prior to the sample peak is selected and merged to obtain an adequate baseline SAXS profile. The user must be more cautious when selecting buffer frames for less optimized experiments to ensure accurate baseline subtraction.

One common example of an improper experimental setup is a linear rise in the baseline of the SEC-SAXS chromatogram (Figure 2-6B). This is typically caused by improperly equilibrating the system before sample injection (See Note 2), or unstable conditions such as temperature causing X-ray optics to move and the beam to shift (See Note 3). Often a reasonable SAXS profile may still be extracted by taking an equal number of frames from before and after the peak for subtraction. If the fluctuations in the SAXS baseline are linear or uniform, it may be possible to correct it mathematically. Recently, programs like REGALS [57] are being utilized for this purpose.

A typical example of a poorly behaved sample (Figure 2-6C), demonstrated as a stepwise rise in the overall background intensity after the main sample peak. This baseline shift commonly occurs from fouling or deposition of sample on the windows of the sample flow cell resulting from an overly concentrated or radiation-sensitive sample. In the case of such an increase in the baseline, which occurs continuously across the elution peak as sample deposits on the flow cell windows, it is often not sufficient to use frames for the baseline subtraction form before or after the sample peak.   Instead, an averaging of an equal number of frames from before and after the peak can produce a more suitable baseline for subtraction and can partially eliminate artifacts of the capillary fouling (Figure 2-6C, D).

Many factors go into determining the quality of the final SAXS profiles but a linear fitting, of low scattering angles, using the Guinier approximation [58],

$$I(q) \ = \ I(0) \exp{(\frac{-q^2 Rg^2}{3})} \hspace{2cm} \text{2-2}$$

with the limit $qRg < 1.3$, provides a quick assessment of the aggregation state of the sample, as established previously [1]. In the case of SEC-SAXS data, the non-linearity of the Guinier region can also indicate the contribution of capillary fouling. A linear Guinier region validates proper buffer subtraction in the final merged data. As example, only when an equal number of buffer frames before and after the sample peak are used in subtraction can artifacts from capillary fouling be eliminated as shown by the linear Guinier region in the inset of Figure 2-6D.
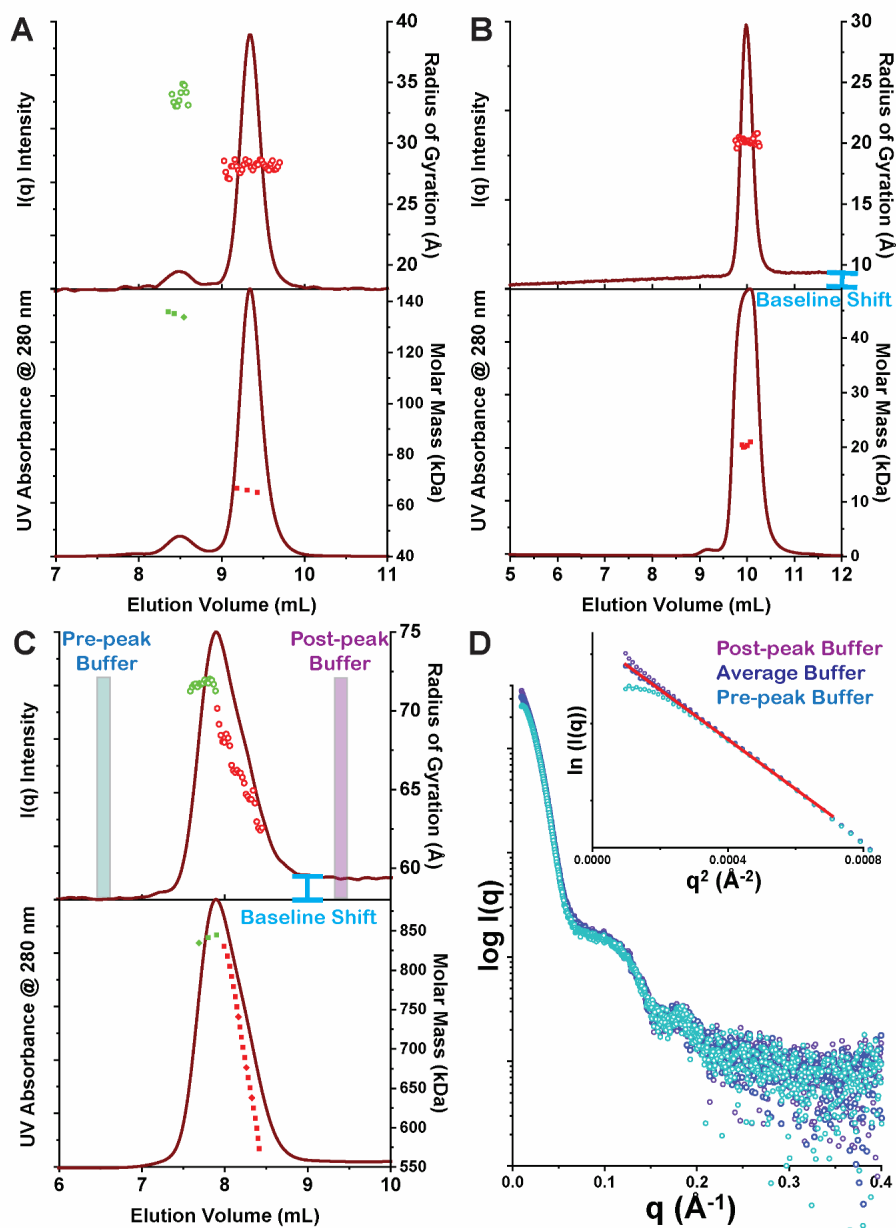
Figure 2-6. Examples of SAXS (top) chromatograms with calculated Rg values and MALS (bottom) chromatograms with calculated MW$_{MALS}$ values for A) a well-behaved sample of bovine serum albumin (BSA) showing dimer and monomer peaks, B) an incorrectly equilibrated sample demonstrating a linearly increasing baseline, and C) a highly concentrated sample showing a stepwise increase in the baseline due to fouling of the flow cell windows. Stability in the protein sample is represented by A) stable Rg and MW$_{MALS}$ values across the peaks, and C) linear change in Rg and MW$_{MALS}$ representing a transient protein complex or confirmation. D) Examples of different buffer regions selected for subtraction to produce SAXS profiles as selected from the stable Rg region in C. Insert shows the linear fit of the Guinier region of each curve plotted as ln $(I(q))$ against q^2.

An estimate of the Rg value, based on the linearity of Guinier region, for each frame across the sample peak can automatically be calculated and plotted by most SEC-SAXS data processing software, including SIBYLS SAXS Process (Figure 2-5 #8). These estimated Rg vs. frame number plot can guide the selection of the best frames across the SEC elution to merge for a final SAXS profile. For well-behaved samples, these estimated Rg values across the sample peak will remain stable (Figure 2-6A, B) merged to produce the highest quality SAXS profile possible for that sample (See Note 4). It is also common to observe fluctuation in Rg values across a single peak, especially when analyzing proteins and protein complexes. These changes in Rg values usually signify a transient behavior of the protein complex. In this case, multiple regions across the peak may be selected and merged independently to produce multiple unique SAXS profiles. This sort of analysis usually constitutes a balancing act between the number of SAXS profiles to produce out of the peak and the number of frames to merge for each one. Several tools are available now to assist in this sort of analysis, like the Evolving Factor Analysis tool [59] included in the RAW [55] software package or Gaussian decomposition as in EFAMIX [54] and US-SOMO [60]. Suppose the frames across the peak cannot be divided due to a low concentration of scattering intensity. In that case, the whole peak can instead be merged and consequently analyzed by multistate modeling [61-63].

2.7.2 Sample Validation Through SEC-MALS

One factor that makes SEC-SAXS-MALS such a powerful technique is that the SAXS data can be validated by the complementary spectroscopy techniques collected in tandem. The UV-vis signal is used to calculate the sample concentration. Still, the wavelengths recorded can be changed to provide additional information like the protein/ nucleic acid ratio [52,64] or the oxidation state of a protein-heme moiety [4]. The UV-vis signal is also typically the most sensitive measurement and the best real-time monitor for the elution of samples during an experiment. The MALS and QELS measurements can provide a wealth of information on their own, most notably MW_{MALS} and Rh, respectively. The MW_{MALS} compliments the MW_{SAXS} derived from SAXS [65] and tends to be more accurate, while the Rh from the QELS is analogous to the Rg from the SAXS with the notable difference that the Rh includes the solvation layer surrounding the macromolecule in the calculation [66]. These complementary in-tandem spectroscopic and light scattering techniques all serve to assess the overall quality of the SAXS measures.

2.7.3 Notes

1. A small, stable buffer region should be defined as a region including about 1-3% of the total frames where the baseline is relatively flat, and no sample is eluting. Typically for well-behaved samples, this region is taken immediately before the sample peak, but immediately after or a combination of both can be taken as circumstances necessitate.
2. For optimal equilibrating of the SEC-SAXS-MALS system, typically equilibration times at 0.65 mL/min are about 90 min when the column is not pre-equilibrated and about 35 min if it is prior to sample injection. The status of the equilibrating is best observed with the MALS and dRI detector following the advised quality checks in the user manuals for those systems (see Section 2.5.5 Note 1).

3. Great care should be taken to minimize vibration and temperature fluctuations which can cause optics to shift the beam to drift. Even movements on the order of 10 µm can be enough to cause baseline fluctuation and aberrant data. This baseline shift is not always apparent in the UV, MALS, or dRI chromatograms (Figure 2-6B).
4. Several factors play into the signal-to-noise ratio in a fully merged SAXS profile, but generally more frames merged across a stable Rg region led to higher signal-to-noise and hence higher data quality.

2.8   Summary and Conclusion

SEC-SAXS-MALS measurements have become widely preferred over traditional SAXS measurements for studying structural biology primarily due to their versatility and *in situ* purification. In this chapter, we first describe the intricacies of the current SEC-SAXS-MALS system design at the SIBYLS beamline which incorporates a general design philosophy of high collection rates and robust automatization. We then discuss the mail-in user program starting with general sample preparation and ending with the use of the SEC-SAXS-MALS data collection, processing, and deposition pipeline made simple through the use of our new SIBYLS SAXS Process beamline control software. Finally, we suggest some general protocols and guidelines for the analysis of SEC-SAXS-MALS data collected at the SIBYLS beamline or elsewhere.

The SEC-SAXS-MALS technique as well as other high-throughput SAXS measurement platforms has been a primary focus of the SIBYLS beamline over the last five years and continues to be a point of active development. As we continue to refine our techniques and streamline our data analysis pipeline, we look to the future of brighter light sources, improved sample delivery options, and advanced computational tools which may be the next revolution in the use of SAXS.

3 Monitoring Nuclease Activity by X-ray Scattering Interferometry using Gold Nanoparticle Conjugated DNA‡

3.1 Chapter Abstract

The biologically critical, exquisite specificity and efficiency of many nucleases, such as those acting in DNA repair and replication, often emerges in the context of multiple other macromolecules. The evolved complexity also makes many biologically relevant nuclease assays challenging and low-throughput. Meiotic recombination 11 homolog 1 (MRE11) is an exemplary nuclease that initiates DNA double-strand breaks (DSB) repair and processes stalled DNA replication forks for restart. Thus, DNA resection by MRE11 nuclease activity is critical for multiple DSB repair pathways as well as in replication. In general, *in vitro* nuclease activity of purified enzymes is studied either through gel-based assays or fluorescence-based assays like fluorescence resonance energy transfer (FRET). However, adapting these methods for a high-throughput application such as inhibitor screening can be challenging. Gel-based approaches are slow and FRET assays can suffer from interference and distance limitations. Here we describe an alternative methodology to monitor nuclease activity by measuring the small angle X-ray scattering (SAXS) interference pattern from gold nanoparticles (AuNPs) conjugated to 5'-ends of dsDNA using X-ray scattering interferometry (XSI). In addition to reporting on the enzyme activity, XSI can provide insight into DNA-protein interactions, aiding in the development of inhibitors that trap enzymes on the DNA substrate. Enabled by efficient access to synchrotron beamlines, sample preparation, and the feasibility of high-throughput XSI data collection and processing pipelines, this method allows for far greater speeds with less sample consumption than conventional SAXS techniques. Importantly due to dominant scattering from AuNPs, the high sensitivity of these XSI assays remains essentially unperturbed by the size or complexity of the protein assembly interacting with the target DNA. The reported metrics and methods can readily be generalized to monitor not only other nucleases but also most other DNA-protein interactions suggesting XSI will prove to be an empowering approach to defining protein-DNA structural interactions.

3.2 Introduction

Meiotic recombination 11 homolog 1 (MRE11) is a critical nuclease that initiates DNA double-strand break (DSB) repair and processes stalled DNA replication forks [67-69]. Thus, DNA resection by MRE11 nuclease activity is important for multiple DSB repair pathways as well as in replication. The crystal structures of MRE11 shows that the active site of MRE11 contains two Mn^{2+} ions and protein form a dimer both in an apo state as well as in DNA-bound states [70-72]. MRE11 is one of the first proteins to respond to DNA damage causing DSBs [67-69]. The DSBs are mainly repaired through either homologous-recombination (HR) that repairs the DSB in an error-free fashion or through non-homologous end-joining (NHEJ) pathway which results in deletions

---

‡ Published as Rosenberg, D. J.; Syed, A.; Tainer, J. A.; Hura, G. L. Monitoring Nuclease Activity by X-Ray Scattering Interferometry Using Gold Nanoparticle-Conjugated DNA. In *DNA Damage Responses*; Mosammaparast, N., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2022; Vol. 2444, pp 183–205.

in the repaired DNA. The first step in the HR repair pathway is MRE11-mediated resection (in the 3'-5' direction) of the DSB leading to 3' ssDNA overhangs [68]. These ssDNA overhangs inhibit NHEJ which requires very little processing of the broken DNA ends. Thus, MRE11 activity is the key determinant of whether DSBs are resected through HR or NHEJ [73].

In general, *in vitro* nuclease activity of purified enzymes is studied either through gel-based assays or fluorescence-based assays like fluorescence resonance energy transfer (FRET). However, adapting these methods for a high-throughput application such as inhibitor screening can be challenging. Gel-based approaches are slow and FRET assays can suffer from interference and distance limitations (~1-10 nm) [74]. In the current method, we are combining our expertise in small angle X-ray scattering (SAXS) with the scattering power of gold nanoparticles (AuNPs) conjugated to dsDNA substrates (dsDNA-AuNP). As will be demonstrated, when AuNPs are held at fixed distances on the 5'-ends of dsDNA, they act as molecular rulers through X-ray scattering interferometry (XSI) and can be used as a high-throughput technique to measure the binding and nuclease activity of MRE11 or other proteins that interact with DNA.

In biological research, SAXS is empowering for structural characterization of biomacromolecules at near physiological conditions [9]. Molecular assemblies, conformational changes, and flexibility can be robustly analyzed from a properly performed SAXS experiment [19]. SAXS is generally performed in solution with modest sample requirements, probing sub-nm distances and microsecond timescales at many synchrotrons around the world. Coupled with sample handling robotics or microfluidics, impactful measurements can be made in high-throughput. Although directional information is lost due to orientational averaging of the macromolecules relative to the probing beam, SAXS provides critical dynamics information to complement atomic resolution techniques like macromolecule X-ray crystallography (MX), nuclear magnetic resonance (NMR) and cryogenic-electron microscopy (cryo-EM) where both the distances and direction between atoms can be recorded [75]. Yet, high-resolution structures from MX or cryo-EM can typically only be attained on specific constructs that are sufficiently homogeneous in conformation and assembly. Fortunately, using these experimental models and SAXS data, models of the full-length or alternate conformations can be determined along with information on flexibility, assembly, and conformational states [76]. SAXS is the right balance between information and throughput for many biological systems [9].

In a typical SAXS experiment, X-ray scattering from the biomacromolecule is measured in a buffer solution. The particle scattering intensity $I(q)$ is a function of momentum transfer $q=(4\pi sin\theta)/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam. $I(q)$ can be derived from the electron distribution within the biomacromolecule as:

$$I(q) = 4\pi \int_0^{D_{max}} P(r) \frac{sin(qr)}{qr} dr \qquad \text{3-1}$$

where r is the distance between election pairs within the macromolecule which leads to a statistical distribution of electron pair distances, or pair-distribution function, P(r), where the maximal dimension, D_{max}, of a molecule is found as the function goes to zero [1,65].

SAXS is inherently a contrast measurement technique where the signal is derived from differences in electron density $\Delta\rho(r)$ between biomolecule $\rho(r)$ and that of the bulk solvent $\rho(s)$ [1] as:

$$\Delta\rho(r) = \rho(r) - \rho(s) \qquad\qquad 3\text{-}2$$

The approximate values for electron density of protein, DNA, and bulk solvent (pure water) are 0.43, 0.55 and 0.33 e⁻/ ų, respectively [77]. Given that the differences in the electron density between the biomolecules of interest and the buffer are already small, very minor fluctuations in buffer composition used for subtraction can greatly affect the results. Thus, a reasonable concentration of the analyte and careful buffer subtraction is essential for obtaining useful information in SAXS experiments. To overcome the challenges in producing the large amounts of protein required for large-scale assays and the sensitivity of buffer fluctuations, high-throughput XSI can be used. This technique expands upon all of the same physical phenomenon of conventional SAXS by utilizing the interference pattern generated not between atom pairs but rather between heavy atom clusters (e.g., AuNPs) held at fixed distances by biomolecules (e.g., dsDNA). These AuNPs function as slits in reciprocal space to the atomic scale wavelengths of hard X-rays in an analogous way as the physical slits in Young's experiments with visible light from classical physics [78]. Importantly, AuNPs having significantly more electron density (5-nm NP ~4.6 e⁻/ ų), scatter X-rays with ~200-fold greater intensity as compared to a 172-kDa protein or 5,400-fold higher than that of a 31-bp dsDNA since scattering intensity on an absolute scale , $I(0)$, (where q=0) is proportional to the square of the number of electrons (m) in a particle [34] as:

$$I(0) = Nm^2(1 - \rho(s)\,\psi)^2 \qquad\qquad 3\text{-}3$$

where N is the number of particles and ψ is ratio between the particle volume and its number of electrons. The original idea of measuring scattering from heavy metals in a biomolecule was proposed in as early as in late 1940s and successfully performed in 1980 [79]. In 2008, Mathew-Fenn et al. were the first group to use dsDNA-AuNP as molecular rulers via XSI [29], applying this technique to measuring the double helix with exceptional accuracy [30]. The P(r) functions derived from these experiments can be divided into two major peak regions. One corresponding to intra-Au and another for the inter-Au distances [30,34,80]. Using this approach, inter-particle distances between two AuNPs separated by up to 100-bp have been accurately measured [80] and greater distances are presumed possible. Other have followed this technique, studying AuNPs conjugated to DNA, RNA, and even proteins using XSI [37,38,81-84].

Our group has applied this XSI technique to probing the mismatch repair of MutS/L via XSI, demonstrating that the technique can be used to study damage-specific structural-changes in the DNA caused by MutS/L [34]. This study focused on the qualitative changes in the inter-particle distances providing information on DNA-protein interactions. In our study of MRE11 nuclease activity we sought to observe such DNA-protein interactions as well as develop a more quantitative assay towards the future of high-throughput XSI experiments. As such we have designed Au-DNA substrates of two different lengths (37-bp and 57-bp) conjugated to 10 nm AuNPs via a Trithiol (TrT) (Letsinger's type) linker on the 5'-end annealed to a shorter (9-bp) duplex forming ssDNA oligo leaving a long stretch of ssDNA available for MRE11 binding. In both substrate cases, the inter-Au distances distributions are shifted to lower mean values compared the substrate alone for samples where nuclease activity was not observed indicating structural changes in the DNA associated with MRE11 binding. As expected, increased MRE11 nuclease activity decreases the population of dsDNA-AuNPs as observed by the decrease in the

amplitude of P(r) corresponding only to inter-Au distances. The intra-Au regions remains unperturbed as the amount of AuNPs is not changing. The P(r) functions are normalized to the intra-Au peak during analysis to account for any minor fluctuations in AuNPs concentration. From these XSI assays, we observe that MRE11 is not active when the active site metal (Mn^{2+}) is not present in the reaction buffer, or a nuclease-dead mutant (H129N) is used in the reaction instead of the wildtype (WT) enzyme.

Prior to analysis, we validated that the substrate is cleaved by MRE11 via gel-based assays using the same substrates as in the XSI experiments except with Fluorescein (6-FAM) substituted for the TrT linkers and AuNPs on the 5'-ends. In general, it is useful (but not essential) to have an independent assay for protein-DNA interactions. For MRE11 the gel data agreed with our XSI activity assays showing that the nuclease activity observed in the reaction with the WT enzyme in the presence of $MnCl_2$. Additionally, these gel-based assays indicate that MRE11 can cut on both strands and on the longer strand it can chew all the way to the 5'-end of the DNA.

Since SAXS probes all molecules in a solution, homogeneous samples are often used [9]. In DNA repair and damage responses there is a need to examine enzyme activities where their active states may be in complexes that are transient and dynamic. To address this challenge, we combined the efficiency of SAXS with the high contrast of Au, and present here a SAXS method with Au-labeled DNA as a robust prototypic assay on DNA processing. More specifically, these experiments can be carried out in any solution condition, in high throughput, provide sub-nm resolution at low concentrations and have the inherent potential to categorize sub-millisecond reaction steps. Furthermore, many DNA repair processes have longer DNA footprints than are comfortably assayed using FRET. This method and the approach defined here for MRE11 can complement and extend more traditional, fluorescent-based assays. High-throughput XSI has a robust ability to test combinations and additives including other macromolecules without loss of signal. These protocols offer strategic and tactical advantages for studies to identify novel inhibitors from screening chemical libraries with the expectation that 1,000 experiments can be done weekly and with batched compounds 10,000 compounds can be screened in one week.

In the following sections we describe our XSI method to conjugate DNA-substrate to the AuNPs for XSI experiments and data analysis protocol. We employ MRE11 as an example but by changing the design of the DNA-substrate, this method can be applied for many other enzymes that are known to cause structural changes in DNA including major types of DNA damage responses. For example, some DNA repair proteins of biological interest bind DNA without making any chemical alterations to control pathway selection [85], however, if these bend DNA or otherwise alter the distance between DNA ends as they typically do, then XSI will provide a sensitive high-throughput measure of their interactions. We therefore expect XSI will be able to interrogate the impacts of proteins and RNA binding to DNA repair and replication complexes, ranging from scaffold proteins such as XRCC1 that is essential to micro-homology mediated end joining [86], to RNA that can act in efficient DSB repair machines [87], to PAR clouds at DNA damage controlled by poly(ADP-ribose) polymerase (PARP1) and poly(ADP-ribose) glycohydrolase (PARG) whose inhibitors are actively being pursued for cancer therapy[88,89], and even to G-quadruplex, repetitive sequence elements, and other non-B DNA sequences associated with DNA instability and mutation sites[90,91]. We demonstrate how to leverage this technique for use at a researcher's home institution as well as how to take advantage of the mail-in user

program of the SIBYLS beamline at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL) helping to design and carry-out experiments like those mentioned herein.

3.3   Materials

1. 15 mg/mL BSPP solution: Dissolve 375 mg bis(p-sulfonatophenyl)phenylphosphine (BSPP) in 25 mL ddH$_2$O.

2. 5 M NaCl solution: Dissolve 146 g sodium chloride (NaCl) in 400 mL ddH$_2$O then add ddH$_2$O until total volume equals 500 mL.

3. 100 mM Phosphate buffer (PBS) pH 7: Dissolve 7.744 g of sodium phosphate dibasic heptahydrate (Na$_2$HPO$_4$•7H$_2$O) and 2.913 g of sodium phosphate monobasic monohydrate (NaH$_2$PO$_4$•H$_2$O) in 400 mL ddH$_2$O. Adjust pH to 7 using HCl or NaOH then add ddH$_2$O until total volume equals 500 mL.

4. Au-BSPP storage buffer (15 mM PBS, 1 mg/ml BSPP, 1 mM TCEP, pH 6.4): Dissolve 50 mg BSPP and 14 mg Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) in 30 mL ddH$_2$O. Add 15 mL 100 mM Phosphate buffer (PBS) pH 7 to solution. Adjust pH to 6.4 using HCl or NaOH then add ddH$_2$O until total volume equals 50 mL.

5. 10nm Au NP: Purchased from Ted Pella

6. 5' Tri-thiolated ssDNA in solution: Purified/lyophilized ssDNA sequences with a Trithiol (Letsinger's type) modification to the 5'-end are purchased from Fidelity Oligos at ~100 nmole scale (Table 3-1. Table showing DNA substrate sequences.) and are re-hydrated in 0.5 mL ddH$_2$O.

7. SH-PEG solution: Thiolated poly(ethylene glycol) (SH-PEG), MW=356.5 was purchased from PolyPure (Oslo, Norway) and 20 μL SH-PEG is added to 480 μL ddH$_2$O.

8. High salt FPLC buffer (15 mM Tris, 1 M NaCl, pH 8): Dissolve 58.44 g NaCl and 1.82 g tris(hydroxymethyl)aminomethane (Tris) in 900 mL ddH$_2$O. Adjust pH to 8 using HCl or NaOH then add ddH$_2$O until total volume equals 1 L.

9. No salt FPLC buffer (15 mM Tris, pH 8): Dissolve 1.82 g Tris in 990 mL ddH$_2$O. Adjust pH to 8 using HCl or NaOH then add ddH$_2$O until total volume equals 1 L.

10. Nuclease reaction buffer for SAXS-based detection (25 mM MOPS, 60 mM KCl, 0.2% T-20, pH 7): Dissolve 2.89 g (3-(N-morpholino)propanesulfonic acid) (MOPS), 2.24 g potassium chloride (KCl), and 1.095 g Tween-20 in 980 mL ddH$_2$O. Adjust pH to 7 using HCl or KOH then add ddH$_2$O until total volume equals 1 L.

11. 20 mM MnCl$_2$ solution: Dissolve 125.8 mg in 50 mL ddH$_2$O.

12. The catalytic domain of human MRE11 nuclease (1-411) selected based on the previous report [72] and is cloned into pET-series expression vector with N-terminus His-tag (Addgene#29653). Surface-exposed methionines (M26, M84, M157, M309, M343) are modified to leucines for improving the protein stability, the modified MRE11 construct maintains the nuclease activity as the parental construct.

13. DH5α chemical competent cells (ThermoFisher)

14. Rosetta^{TM} chemical competent cells (Novagen)

15. BD Difco^{TM} LB Broth, Miller (Lauria-Bertani) media

16. BD Difco^{TM} LB Agar, Miller (Lauria-Bertani) media

17. Kanamycin sulfate UPS grade (Teknova)

18. Lysis buffer (50 mM Tris (pH-7.5), 500 mM KCl, 5% Glycerol, 0.5% T-20, Protease Inhibitors)

19. Buffer A (25 mM Tris (pH-7.5), 300 mM NaCl, 2.5% Glycerol, 20 mM Imidazole)

20. Buffer B (25 mM Tris (pH-7.5), 300 mM NaCl, 2.5% Glycerol, 500 mM Imidazole)

21. SEC buffer (20 mM Tris(pH-8.0), 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT)

22. HisTrap FF Crude pre-packed 5 mL column (GE/Cytiva)

23. Hi Load ^{TM} 16/600 Superdex200 pg (GE/Cytiva)

24. Nuclease reaction buffer for FAM-based detection (25 mM HEPES (pH=7), 50 mM KCl, ±1 mM $MnCl_2$)

25. To make 3X Stop Buffer mix 0.5 mL formamide, 0.12 mL of 0.5 M EDTA (pH=8), 0.25 mL of 100% Glycerol, and 0.15 mL of 10% SDS.

26. 4-20% Mini-PROTEAN TGX Stain-Free protein gels (Bio-rad)

27. 15% Mini-PROTEAN TBE-Urea gel (Bio-rad)

28. DNA Sequences (from IDT) used in the gel-based assay are given in Table 3-1:

Table 3-1. Table showing DNA substrate sequences.

*37-bp-Au-F	5'-FAM-TTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-3' 3'-CGGCCCGCGT-5'
*37-bp-Au-R	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-3' 3'-CGGCCCGCGT-5'-FAM
*57-bp-Au-F	FAM-5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-3' 3'-CGGCCCGCGT-5'
*57-bp-Au-R	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-5' 3'-CGGCCCGCGT-5'-FAM
**37-bp-XSI Substrate	5'-(Au-NP)-TrT-TTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-3' 3'-CGGCCCGCG-TrT-(Au-NP)-5'
**57-bp-XSI Substrate	5'-(Au-NP)-TrT-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCGGGCGC-3' 3'-CGGCCCGCG-TrT-(Au-NP)-5'
*FAM = Fluorescein (6-FAM), **TrT = Trithiol linker (Letsinger's type)	

3.4 Methods

3.4.1 Preparation of BSPP Protected Au Nanoparticles via BSPP-Citrate Exchange (Au-BSPP)

1. Add 25 mL of 15 mg/mL BSPP to 400mL of either purchased 10nm colloidal AuNPs and filter solution through 0.22 um filter.

2. Stir 400 mL of citrate stabilized colloidal AuNPs with BSPP overnight.

3. Add 5 M NaCl until the solution turns from red to dark red/purple (~75 mL).

4. Pour into Beckman 100 mL polypropylene bottles w/cap assembly.

5. Spin in Beckman centrifuge in JA-18 rotor @ 12000 G for 10mins.

6. Decant slowly or pipette off supernatant (*see* Note 1).

7. Use 0.5 M NaCl solution to wash AuNPs, sonicate, and repeat step 6.

8. Repeat step 7 twice.

9. Resuspend in 25 mL Au-BSPP storage buffer.

3.4.2 Au-ssDNA Conjugation, Anion Exchange Chromatography, and dsDNA-AuNP Annealing

1. One limitation of Au NP conjugation to DNA is feasible only in the 5' end of the DNA, thus we designed a substrate (Figure 3-1, Table 3-1) that would be cut by MRE11 as well as leads to separation of paired AuNPs upon the nuclease reaction.

2. For the conjugation, measure the concentrations of Au-BSPP and ssDNA solutions (diluted appropriately *see* Note 2) using Thermo Scientific NanoDrop 2000 Spectrophotometer at 520 nm and 260 nm respectively.

3. Calculate concentration using Beer's law and the appropriate extinction coefficients for AuNPs and ssDNA (*see* calculation Note 3).

4. Colloidal Au-BSPP and selected ssDNA solutions are mixed at a mole ratio of 3:1 and shaken gently at room temperature (RT) overnight.

5. SH-PEG solution is added to final to mixture at v/v% ratio of 10% (i.e., 100 μL added to 1000 μL solution) and mixture is shaken gently at RT for 2 h.

6. Separate and collected mono-conjugated Au-ssDNA from multi-conjugated using a Dionex DNA-Pac PA100 anion exchange column on an AKTA series fast protein liquid chromatography (FPLC) (*see* FPLC Method Note 4, Figure 3-1A, B).

7. Complementary Au-ssDNA conjugates are annealed by heating at 94 C for 3 min and allowing to cool to RT slowly to form final dsDNA-AuNP substrates (Figure 3-1C).

8. Final dsDNA-AuNP substrates are observed via XSI to ensure inter-particle signal only seen from the properly annealed substrate (Figure 3-1C).
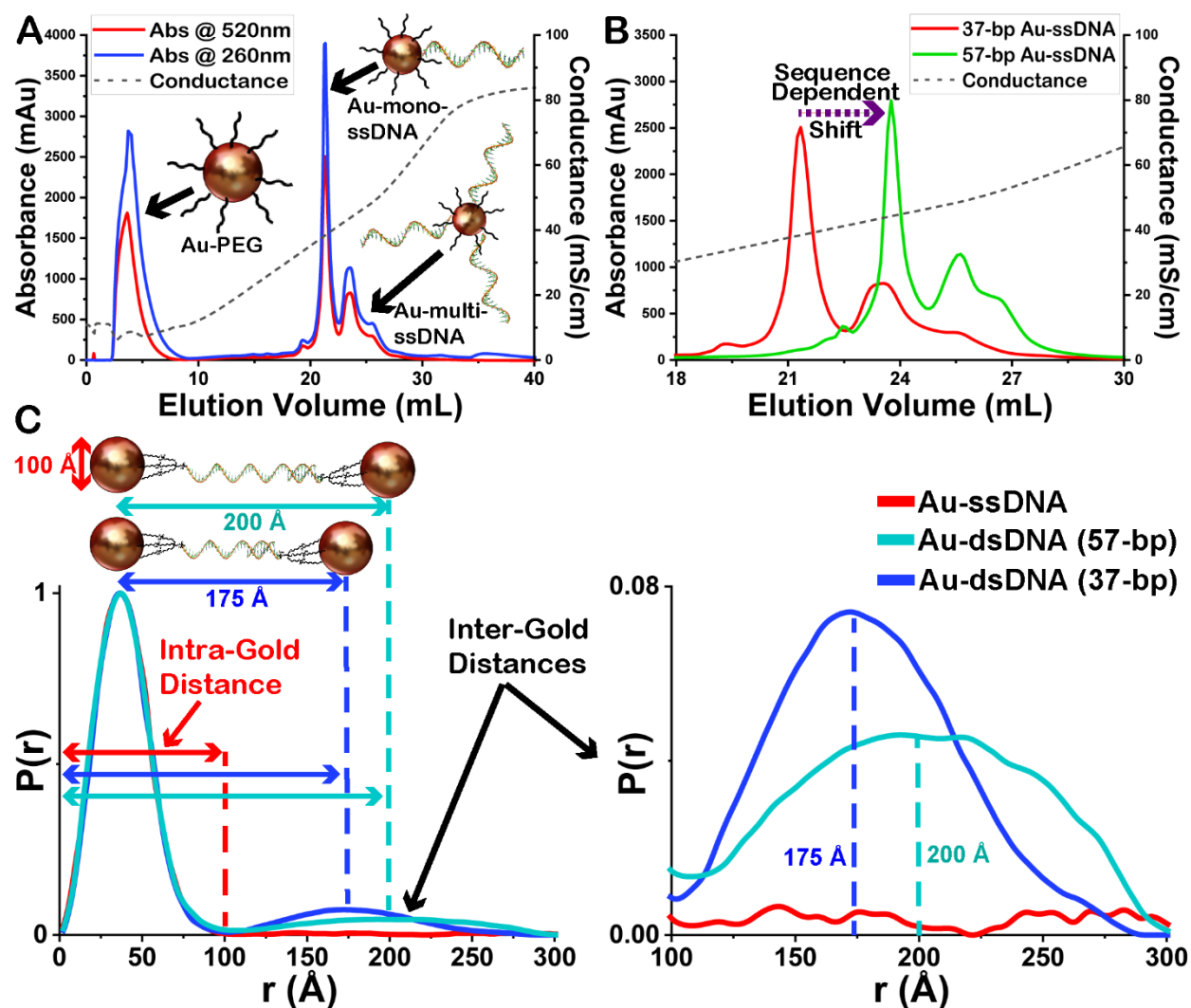
Figure 3-1. Anion exchange chromatograms showing A) the separation of mono-conjugated Au-ssDNA from un-conjugated Au-PEG and multi-conjugated Au-ssDNA using fast protein liquid chromatography (FPLC), and B) the sequence dependent shift in the elution volume. Both as measured by diode array detector (DAD) at 520 nm and 280 nm for AuNPs and ssDNA respectively as well as an additionally, measurement of conductance showing salt gradient conditions. C) Demonstration of the two dsDNA-AuNP substrates used and the normalized electron-pair distance distribution P(r) functions from these experiments showing the peak regions corresponding to the intra-Au and inter-Au distances as well as the disappearance of the inter-Au distances in the Au-ssDNA sample. The P(r) functions are normalized to the intra-Au peak to compensate for fluctuations in concentration (Au NP conc 200 nM ± 10).

3.4.3 Protein Expression and Purification

1. After expression plasmids are verified through DNA sequencing, plasmids are amplified by transforming into DH5α cells and cells are grown on LB-agar plates with Kanamycin

selection (50 μg/mL) overnight at 37 °C and are extracted using Qiagen® miniprep kit as per the manufacturer protocol.

2. For protein expression, extracted plasmids are transformed into Rosetta™ competent cells in a similar fashion as above (*see* Note 5).

3. Expression-plasmid transformed Rosetta™ cells are inoculated into a small LB culture media (200 mL) supplemented with Kanamycin (50 μg/mL) and grown overnight at 37 °C in a shaker.

4. Overnight culture is further utilized to inoculate large-scale (6L) LB media (1.5 L/flask) supplemented with Kanamycin (50 μg/mL) and protein expression is induced with 0.75 mM IPTG at 16 °C overnight.

5. Cells are harvested and stored in -80 °C deep freezer until further use.

6. Cell pellets are thawed and resuspended in the lysis buffer and homogenized using a Dounce homogenizer.

7. Homogenized cells are lysed by sonication.

8. Lysed cells are clarified by centrifugation at 18000 rpm for 45 minutes.

9. Clarified lysate is loaded onto a prepacked 5 mL HisTrap column (pre-equilibrated with Buffer A) mounted on an FPLC system (e.g., AKTA Pure) for an automated affinity purification (column wash-100 mL of Buffer A, second wash-25 mL of 10% Buffer B and elution-50 mL of 60% Buffer B, second elution- 50 mL of 100% Buffer B). Protein eluted with 60% Buffer B is used for downstream activity assays.

10. Eluted protein fractions are verified by protein gel-electrophoresis and protein containing fractions are pooled and concentrated and loaded onto pre-equilibrated (with SEC buffer) Superdex 200 16/600 column mounted on an AKTA pure machine for further purification by size-exclusion chromatography.

11. Protein fractions are verified by gel-electrophoresis and protein containing fractions are pooled and concentrated and quantified by NanoDrop.

12. Protein is distributed into 20-30 μL fractions and flash frozen in the liquid nitrogen and stored in -80 °C deep freezer until further use.

13. Plasmid for the nuclease-dead version of the enzyme (H129N) was generated through mutagenesis and purified exactly as the wildtype (WT) enzyme.

14. Given the composition of SEC buffer contains 0.1 mM EDTA, the purified proteins at the end are in a metal-free state.

3.4.4   DNA Substrate Preparation for the Fluorescence-based Nuclease Reaction

1. Identical DNA substrates are used in Au-SAXS and gel-based nuclease reactions (*see* Note 6 and Table 3-1).
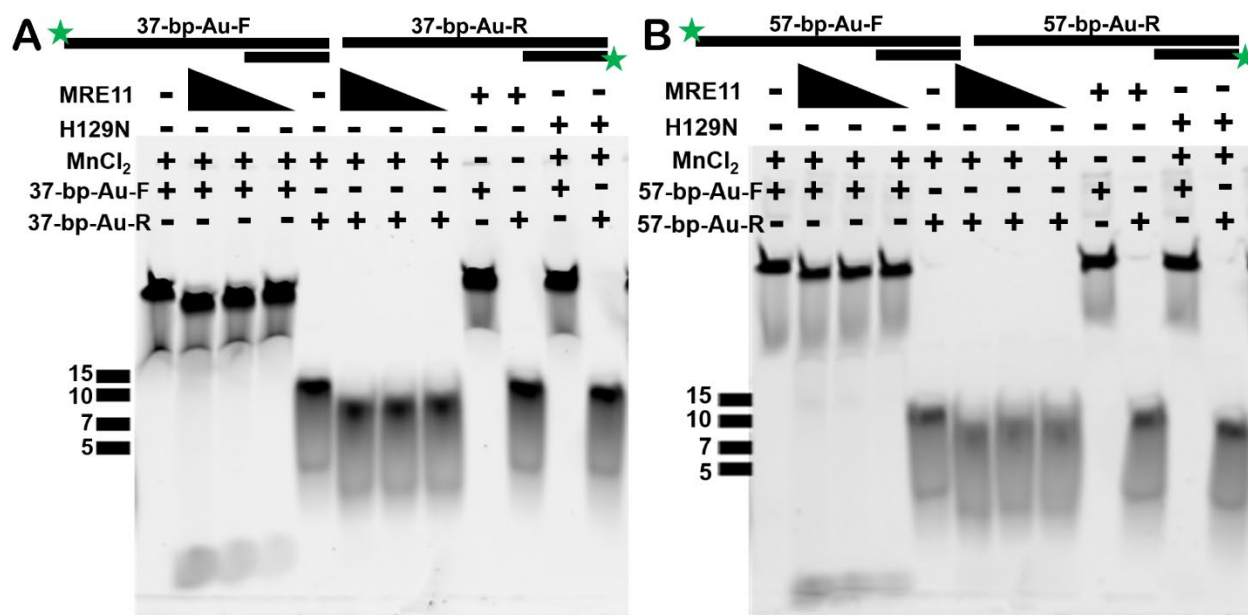
Figure 3-2. MRE11 nuclease activity as monitored in gel-based assays. Both substrates (37-bp and 57-bp) used in the XSI experiments were used in fluorescence-based nuclease assay. To monitor the nuclease activity of MRE11 on both strands of the duplex substrate, a 5'-FAM label is added on either end resulting in four substrates as shown above (37-bp-Au-F, 37-bp-Au-R, 57-bp-Au-F and 57-bp-Au-R). A) MRE11 shows nuclease activity (at 2, 1 and 0.5 μM concentration) on both strands of 37-bp substrate and the activity is dependent on the presence of MnCl$_2$ in the reaction buffer. As expected, the nuclease dead mutant H129N is not active even in the presence of MnCl$_2$ at 2 μM enzyme concentration. B) MRE11 shows nuclease activity (at 2, 1 and 0.5 μM concentration) on both strands of 57-bp substrate and the activity is dependent on the presence of MnCl$_2$ in the reaction buffer. As expected, the nuclease dead mutant H129N is not active even in the presence of MnCl$_2$ at 2 μM enzyme concentration. ssDNA markers are indicated for each gel.

2. All 5'-Fluorescein (FAM) labelled DNA oligos are purchased from IDT with HPLC purification.

3. We verified that MRE11 cuts our substrate thorough monitoring the cleavage in a fluorescence-based nuclease assay (Figure 3-2).

4. To monitor how MRE11 cuts the DNA substrates on both strands, both 37-bp and 57-bp substrates are labelled with FAM at 5' individually resulting in four different substrates: (1) duplex with a 5'-FAM on longer strand of 37-bp, (2) duplex with 5'-FAM on shorter strand of 37-bp, (3) duplex with a 5'-FAM on longer strand of 57-bp and (4) duplex with a 5'-FAM on shorter strand of 57-bp.

5. DNA substrates (in Table 3-1) used in gel-based nuclease reaction are prepared by annealing complementary non-labelled strand with fluorescently labelled oligo (in 1.3:1 ratio) and by

heating at 95 °C for 5 minutes followed by gradual cooling to room temperature for the duplex formation.

6. Substrates are stored at -20 °C until further use (@ 1 µM stock concentration)

3.4.5   Fluorescence-based Nuclease Reaction to Validate Substrates and the Activity

1. Proteins (WT or H129N) are diluted to the desired concentration in the nuclease reaction buffer with or without $MnCl_2$.

2. Nuclease reaction is initiated by adding the substrate to the reaction mixture and incubating at 37 °C for 1 hr.

3. Nuclease reaction is stopped by adding a stop buffer and incubated further at 37 °C for 15 minutes (*see* Note 7).

4. For each substrate, a non-labelled version of the cleaved FAM-labelled strand is added (100-200-fold excess) to the reaction mixture to visualize only FAM-labelled ssDNA product.

5. Reaction mixture is run on a denaturing TBE-UREA gel for 50 minutes at 185 V (*see* Note 8).

6. Gel can be imaged with FAM excitation/emission filter on any gel imager (Figure 3-2).

3.4.6   Sample Preparation of XSI Experiments

1. Dialyze dsDNA-AuNP substrates overnight at 4 °C in 1 L reaction buffer using 4 kDa dialysis membranes. Be cautious of strong reducing agents in the buffer (*see* Note 9).

2. Measure the concentrations of dsDNA-AuNP (diluted appropriately *see* Note 2) using Thermo Scientific NanoDrop 2000 Spectrophotometer at 520 nm. Adjust concentration if needed (*see* Note 10).

3. Combine enzymes with dsDNA-AuNP in an Axygen 96-well Polypropylene PCR Microplate at a final molar ratio of 10:1 (MRE11 2 μM and AuNPs 200 nM) in Nuclease reaction buffer (with or without 2 mM $MnCl_2$) and then bracket the samples with a blank buffer sample on either end for buffer subtraction (*see* Note 11, Figure 3-3).

4. Incubate plate containing samples at 37 °C for 1 h for the reaction to take place.

3.4.7   XSI Data Collection at the SIBYLS Beamline (see Note 12)

1. XSI data is collected at the SIBYLS beamline (BL12.3.1), at the Advanced Light Source at Lawrence Berkeley National Laboratory, Berkeley, California [2]. To send samples for collection *see* Note 12.

2. Load 96-well sample plate onto cooled 10 °C sampling position.

3. Samples are transferred from a 96-well plate via a Tecan Evo liquid handling robot with modified pipetting needles acting as sample cells to the X-ray beam as described previously [9].

4. X-ray wavelength is set at $\lambda = 1.24\ Å$ and the sample-to-detector distance is 2.1 m, resulting in scattering vector q, ranging from 0.01 Å $^{-1}$ to 0.45 Å $^{-1}$. The scattering vector is defined as $q = 4\pi sin\theta/\lambda$, where 2θ is the scattering angle. Data is collected using a Dectris PILATUS3X 2M detector at 20°C and processed as previously described [10]. Samples are exposed to X-ray synchrotron radiation for a total of 10 seconds at a frame rate of 0.2 seconds for a total of 50 images.

5. For each sample collected, two sample-free buffer samples are also collected to reduce error in subtraction (Figure 3-3).

6. Each collected image is circularly integrated and normalized for beam intensity to generate a one-dimensional scattering profile by beamline specific software (Figure 3-3).

7. Buffer subtraction is performed for the one-dimensional scattering profile of each sample by using each of the two corresponding buffers, producing two sets of buffer-subtracted sample profiles to ensure the subtraction process was not subject to instrument variations (Figure 3-3).

3.4.8 Setting up XSI Data Processing Pipeline

1. Once data collection has been completed you will receive your data back with the following file hierarchy:

```
Username_Date:
├── Results
├── Subtracted
│   └── A2_results
│       ├── Average
│       │   └── 50 Integrated Curves (.dat files)
│       ├── Buffer1
│       │   └── 50 Integrated Curves (.dat files)
│       └── Buffer2
│           └── 50 Integrated Curves (.dat files)
└── Unsubtracted
    ├── A1b_results
    ├── A2_results
    └── A3b_results
```
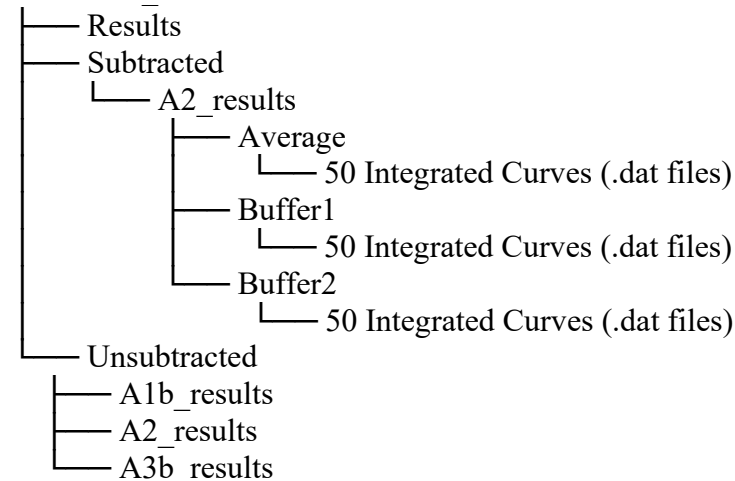
2. Scattering profiles over the 10 second exposure (50 frames total) should be sequentially averaged to eliminate any radiation damage affects. This can be done either manually, for each sample using our web-based beamline software FrameSlice (sibyls.als.lbl.gov/ran), or by batch processing using our XSI data processing pipeline (Figure 3-3) which is recommended for large data sets.
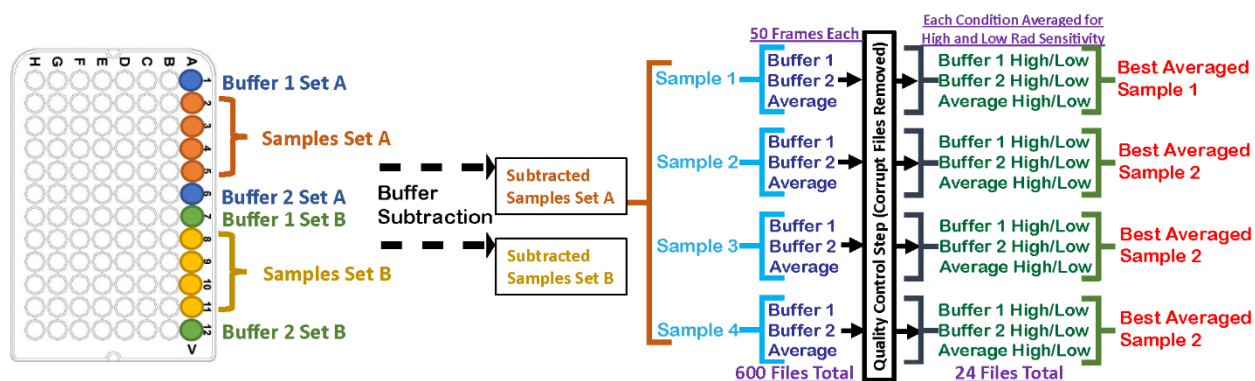
Figure 3-3. Exemplary demonstration of how to setup a 96-well plate for high-throughput XSI assay and the subsequent data processing pipeline.

3. To set up your system for running the frame averaging pipeline we have recommended Bash terminal environment (*see* Note 13), but it should be able to run on any platform with Python 3 and pip (both required).

4. Check your versions of python and pip.

To check your python version from terminal:

$ python3 --version

If there is no version of python 3 (*see* Note 13).

To check your pip version from terminal:

$ python3 -m pip --version

If no version of pip type:

$ python3 get-pip.py

5. In a new bash terminal clone our gitlab repository (*see* Note 14):

$ git clone https://git.bl1231.als.lbl.gov/djrosenberg/frame_averaging_pipeline.git

6. Go to the folder called frame_averaging_pipeline and install:

$ cd frame_averaging_pipeline

$ pip install .

To make sure pip has installed frame_averaging_pipeline:

$ pip list

Note the location of the repository folder, frame_averaging_pipeline:

$ pwd

Example Output: *folder_path*/frame_averaging_pipeline

This output we will call *folder_path* (needed in Step 7 to run the main script)

7. Start Xserver if on Windows or Mac and leave it running in the background (*see* Note 15).

8. If your data is local, *cd* to the folder containing the *Results* folder you would like to process (called *Username_Date* in the file hierarchy example above) and run the *xsi_batch_processing* shell script.

Here we use the example of the *Test_Data* included in the *frame_averaging_pipeline* folder:

$ cd tests/Test_Data

Run *xsi_batch_processing.sh* using the path from Step 4:

$ *folder_path*/frame_averaging_pipeline /xsi_batch_processing.sh

9. When asked "Is your data on your local machine and are you in the folder containing your Results folder" answer "y" or "yes"

10. The data processing pipeline should start. Once complete you are asked to "Please Review Output In" the folder. Scroll through the .png images in the Xserver window or preferred image viewer if prompted (*see* Note 15) and decide whether buffer subtraction one, two, or the average should be used and enter 1, 2, or A respectively (if buffers match closely use average). Then select whether the samples show High or Low sensitivity to radiation and enter either H or L respectively (see Figure 3-4).

11. Make sure the desired data has been selected and the output directory is correct and enter "Y" to continue or "N" to repeat selection step 9.

12. To pull your data directly from the SIBYLS beamline database (must have an account, *see* Note 12), answer "n" or "no" when asked "Is your data on your local machine and are you in the folder containing your Results folder".

13. When asked, "Please Enter Your SIBYLS Username: (this is caps sensitive)"
14. When asked, "Is *Current_Year* the correct year of your data collection?" answer "Y" to continue or "N" to enter the year of your data collection as YYYY.
15. Enter user *password.*
16. When asked, "Please Select Data Folder Name You'd Like to Work on (this is caps sensitive) (For example: 2020_02_25_*username*_results)"
17. Make sure the folder and path are correct and answer "Y" to continue or "N" to repeat data selection step 14.
18. Enter user *password*.
19. Batch data processing will start automatically. Follow steps 9 and 10.
20. Once XSI batch processing is completed, the following file hierarchy is output:

Username_Date:

```
├── Best_Curves                    (best selected averaged curves)
├── Results                        (original subtracted/unsubtracted data)
├── Username_Date _Ave_Curves
│   ├── Full_Ave                   (full frame averaged curves (low radiation))
│   └── Region_Ave                 (Regionally averaged curves (high radiation))
├── Username_Date _Ave_Graphs      (.png images of averaged curves)
└── logs                           (logs for debugging code)
```
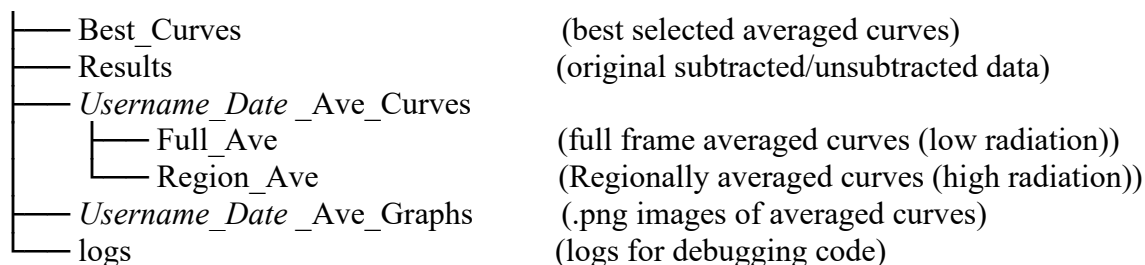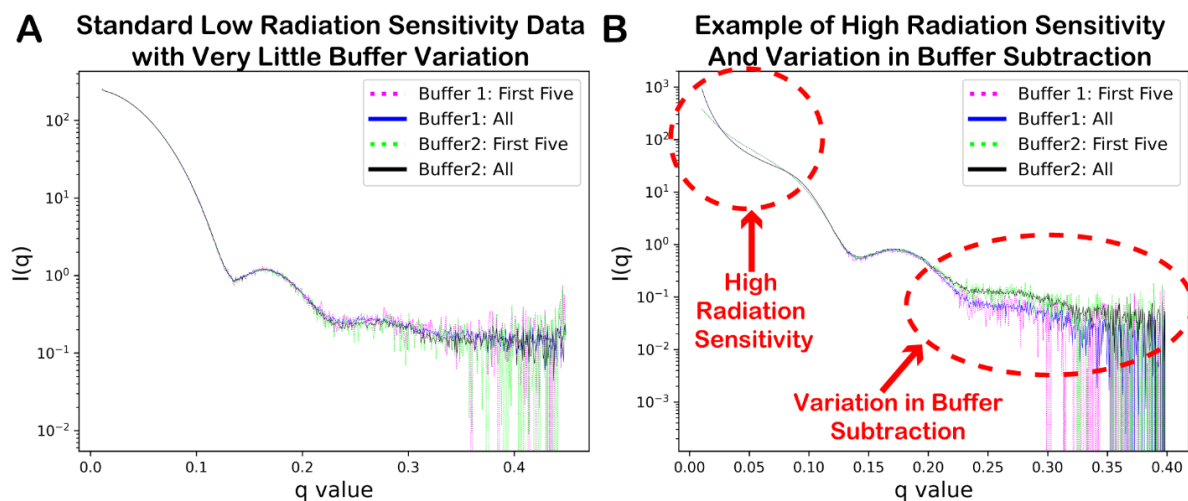
Figure 3-4. Output from the xsi_batch_processing.sh script with examples of A) standard low radiation sensitivity data, and B) high radiation sensitivity with variations in buffer subtraction.

3.4.9 XSI Data Analysis and Interpretation

1. Once you have your best frame averaged XSI curves (*best_curves* in the above file hierarchy) you can use those for analysis.

2. The simplest analysis is to generate pair-distribution functions, P(r), from the inverse Fourier transformation of the best averaged XSI profiles [1,33]. For this we recommend using SCÅTTER [54] (*see* Note 17) but there are many other options [55,92].

3. P(r) functions were normalized to the intra-Au peak to account for variations in concentration.

4. A shift in the peak maximum in the inter-Au region indicates a highly accurate change in inter-particle distance suggesting a change in the substrate. For the example of MRE11, the inter-Au distances distributions are shifted to lower mean values compared the substrate alone for samples where nuclease activity was not observed indicating structural changes in the DNA associated with MRE11 binding (Figure 3-5). These findings are consistent with both dsDNA-AuNP substrate lengths.

5. The integration of the inter-Au peaks is used to estimate the relative changes in concentration for intact dsDNA-AuNP substrates after the enzymatic reaction takes place. As expected, increased MRE11 nuclease activity to decrease in the population dsDNA-AuNP as observed by the decrease in the amplitude of P(r) corresponding only to inter-Au distances. From these XSI assays, we observe that MRE11 is not active when the active site metal (Mn^{2+}) is not present in the reaction buffer or a nuclease-dead mutant (H129N) is used in the reaction instead of the wildtype (WT) enzyme (Figure 3-5). These findings are consistent with both dsDNA-AuNP substrate lengths.

39
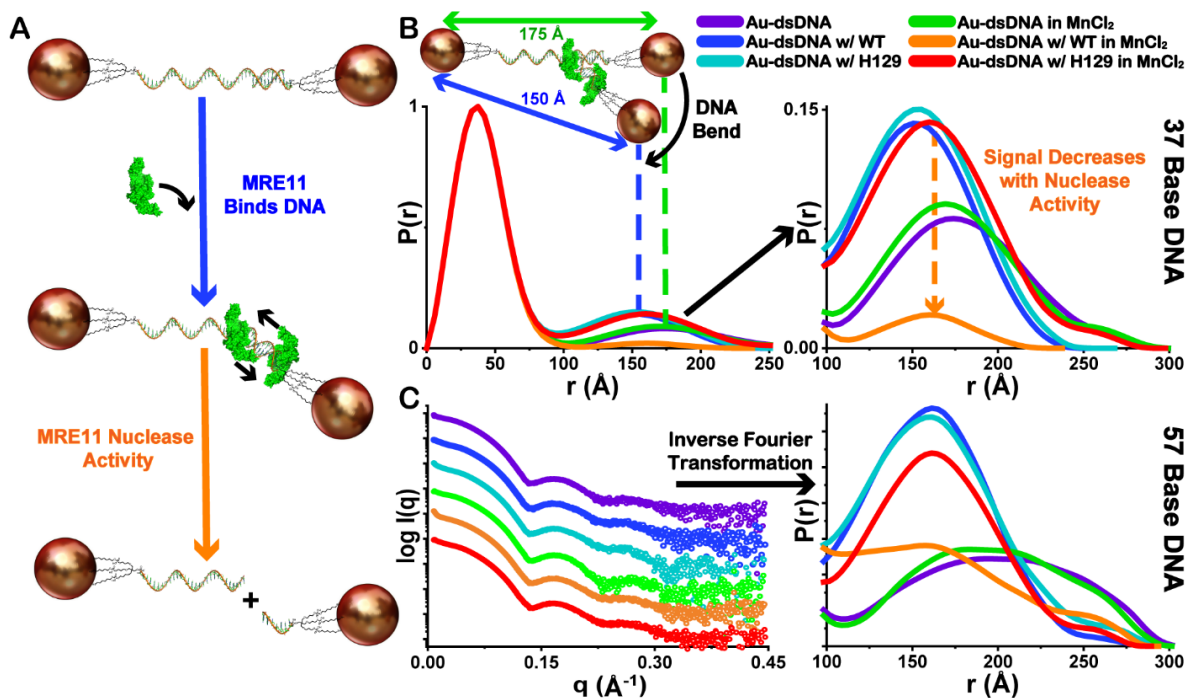
Figure 3-5. Demonstration of overall XSI assay scheme. A) the proposed mechanism of MRE11 interaction with intact dsDNA-AuNP substrates and the subsequent nuclease activity leading to separation of the fixed inter-particle distances as Au-ssDNA. B) Demonstration of the shifts in the distribution of inter-Au electron-pair distances, seen in the normalized P(r) functions (37-bp DNA), to lower mean values compared the substrate alone representing the structural changes in the dsDNA-AuNP substrates associated with MRE11 binding. Additionally, a decrease in the amplitude of P(r) corresponding only in the inter-Au regions is observed only with WT-MRE11 in the presence of $MnCl_2$ (orange) suggesting increased MRE11 nuclease activity (dsDNA-AuNP to Au-ssDNA). Legend for sample identity shown. C) Exemplary experimental XSI curves and derived P(r) functions for 57-bp DNA colored as in the Panel B. Curves have scaled I(q) for visualization purposes. P(r) function plot is scaled to shown inter-Au distance region. All P(r) functions are normalized to the intra-Au peak to compensate for fluctuations in concentration as depicted in Figure 3-1C.

3.5   Notes

1. Au NP pellets can be disturbed easily if not careful. It is recommended to decant with particles held at side closest to the floor if decanting or to use a pipette.

2. To measure concentrations accurately within the confines of Beer's law, analytes should be diluted so that the absorbance range is between 0.1 and 1 mAu.

3. Extinction Coefficient Au @ 520 nm = 9.69×10^6 L mol^{-1} cm^{-1}, Path Length Nanodrop = 0.1 cm, Extinction Coefficient ssDNA @ 260 nm sequence dependent.

$$\frac{Absorbance * Dilution\ Factor}{Path\ Length * Extinction\ Coefficient} = Concentration$$

4. Using no salt and high salt FPLC buffers a salt gradient from 10 mM to 1000 mM is created over a period of 50 minutes (Figure 3-1A, B). Sample elution monitored UV-Vis absorption at the Au plasmon maximum of 520 nm. Typical final concentrations for collected conjugates were 0.1 to 0.2 μM.

5. The catalytic domain of MRE11 is not super soluble when expressed in E coli, however, soluble fraction of the purified enzyme is active in the nuclease reaction.

6. Prior knowledge of enzyme-DNA substrate reaction can be quite useful in designing the substrates. By carefully modifying the substrate this method can be adapted for other DNA nucleases.

7. Enzymes can be removed from the substrate if desired by adding Proteinase K at the end of the reaction.

8. It is recommended to pre-run the TBE-Urea gel (@200V for 60 minutes) prior to running the nuclease reaction products on the gel.

9. DNA is conjugated to AuNPs through Au-S interaction. Strong reducing agents in the reaction buffer can disrupt this interaction and cause aggregation of the AuNPs, especially when exposed to strong synchrotron X-ray radiation. Thus, the reaction conditions need to be optimized accordingly.

10. Samples can be diluted with reaction buffer or concentrated by using 4 kDa centrifuge concentrator tubes and spinning at 10000 G. Generally, concentrations > 100 nM give great scattering signal.

11. SAXS is a contrast measurement, so the buffer used for subtraction must be as close to the buffer containing the sample as possible. Dialysis samples and use the dialysis buffer for best subtraction.

12. While it should be possible to leverage this technique for use at a researcher's home institution, we also offer a mail-in user program at the SIBYLS beamline where we can help to design and carry-out experiments like those mentioned herein. To obtain XSI data collection time at the SIBYLS beamline please follow the directions on our website https://bl1231.als.lbl.gov/htsaxs/instructions/htsaxs and/or contact us.

13. The use of Python 3 is required. The desired version of Python 3 can be installed by follow the instruction on https://www.python.org/downloads/ and any version should work. We also recommend running bash terminal in a Conda, Python 3.7+ environment as it may streamline the setup of the code, but Conda is not required. To setup your own conda environment follow the instructions on https://docs.conda.io/projects/conda/en/latest/user-guide/install/index.html for setting up miniconda on your system if desired.

14. This code under active development and the newest setup and usage information can be found in the README file at our gitlab for frame_averaging_pipeline at: https://git.bl1231.als.lbl.gov/djrosenberg/frame_averaging_pipeline.git

15. If running xsi_bath_processing.sh on Windows or Mac, you will likely need an Xserver (https://kb.thayer.dartmouth.edu/article/336-x11-for-windows-and-mac) to run graphical

interfaces and you will also need Eye of Gnome (eog) installed in your terminal in you don't already have it (*see* Note 16). Alternatively, just open the .png images in your preferred image viewer and *ignore* Notes 15 and 16.

16. The xsi_bath_processing.sh will prompt you if "Eye of Gnome (eog) could not be found. Please use your preferred image viewer to view .png files in the "*username_date_*Ave_Graphs" folder. To install eog:

$ eog --version   # first check if you have eog installed

$ sudo apt-get install eog   #if you don't have it (ubuntu)

$ sudo yum install eog   #if you don't have it (centos/redhat)

17. The SAXS analysis software SCÅTTER can be downloaded from http://www.bioisis.net/tutorials and the website includes tutorials for its use.

4 Defining the Molecular Basis of PARP-1 Damage Recognition, Activation, and Inhibition During Single-Stranded DNA Break Repair Through X-ray Scattering Interferometry

4.1 Chapter Abstract

Due to the clinic importance of poly(ADP-ribose) polymerase 1 (PARP-1), a great deal of effort has gone into investigating its damage recognition, activation, and inhibition during DNA double-strand break (DSB) and single-strand break (SSB) repair. The highly modular nature of the PARP-1 architecture has historically impeded structural studies in solution, thus new techniques must be developed. Herein we study PARP-1 during SSB repair to elucidate the molecular basis behind PARP-1 damage recognition and activation and define the allosteric changes responsible for variations in PARP-1 inhibition. We create gold nanoparticle conjugated double-stranded DNA substrates (dsDNA-AuNPs), to use as molecular rulers via a high-throughput x-ray scattering interferometry (HT-XSI) data collection, processing, and analysis pipeline. This work provides insight into PARP-1 biochemistry during SSB repair, in solution, with nanometer precision previously unobtainable by conventional techniques.

4.2 Introduction

Poly(adenosine diphosphate (ADP)-ribose) polymerases (PARPs) are found acting in nearly any process involving DNA. PARPs catalyze the transfer of ADP-ribose from nicotinamide adenine dinucleotide (NAD$^+$) and are classified as diphtheria toxin type ADP-ribosyltransferases domain (ARTDs) proteins[93]. The first and most extensively explored PARP to be identified in humans is PARP-1, best known for its involvement in SSB[94-99] and DSB[100,101] repair pathways as well as stabilizing DNA replication forks[102], modulating transcription[103], and modifying the structure of chromatin[104]. Due to the ubiquitous nature of PARP-1 in cellular processes, there has been a growing interest in the development of chemotherapeutics targeting PARP-1 over the past decade. This research has led to the development of successful cancer treatments based on PARP inhibition giving rise to synthetic lethality in tumor cells[105-107]. In addition to cancer therapy, evidence suggests that modulating the activity of PARP-1 may also be effective in the treatment of cardiovascular disease[89], Cockayne Syndrome, Xeroderma Pigmentosum A[108,109] and neurodegenerative diseases[110,111]. Even though PARP-1 was the first of the PARP family to be identified, the highly modular nature of PARP-1's "beads-on-a-string" architecture[101,112] has historically impeded structural studies of the full-length protein but is likely responsible for its ability to achieve its diversity of cellular functions. Thus, much of the molecular basis behind damage recognition, activation, and inhibition of PARP-1 is yet unknown. This is especially true in the context of its involvement in SSB repair, inhibition of which is key to attaining successful synthetic lethality in cells lacking a functional DSB repair pathway[105,106]. All PARP inhibitors currently available or in clinical trials work by competitively binding to the NAD$^+$ catalytic active site, however they all exhibit vastly different inhibitory potency[89]. Understanding variations in the mechanism of action between these small molecules would provide a means to tune PARP-1 inhibition and develop new chemotherapeutics.

The basic architecture of PARP-1 involves six domains arranged as "beads-on-a-string" (Figure 4-1A)[101,112]. The first three are zinc-finger binding domains; Zn1 and Zn2 N-terminus domains (PDB: 2dmj and 2cs2, respectively)[95,96,113] containing recognition sites for non-specific DNA binding[95,96,113], and Zn3 (PDB: 2jvn)[114,115] which contains essential residues for interdomain contact involved in domain positioning during DNA-binding (Figure 4-1B,C)[114,115]. Next is the automodification domain (AD) which contains the majority of acceptor residues for the covalent attachment of poly(ADP-ribose) (PAR) during poly(ADP-ribosyl)ation (PARsylation)[116]. The AD also contains a breast cancer 1 protein (BRCA1) C-terminus (BRCT) fold (PDB: 2cok)[117] which has proven non-essential to DNA-binding during SSB or DSB repair[117] but may be involved in the controversial homodimerization of PARP-1[118]. After that comes the WGR domain (PDB: 2cr9), an 80-90 amino acid segment named for its richness in tryptophan (W), glycine (G), arginine (R) residues, known to be essential to interdomain communication and DNA-binding[119]. Additionally, the WGR domain has proven to be involved in the DNA-dependent release of PARP-1 allowing it to search DNA near the diffusion-limited rate[120,121]. Finally in the chain comes the catalytic domain (CAT) (PDB: 1a26)[122], which is composed of the helical subdomain (HD) containing a leucine-switch critical to DNA-dependent activation[123] and the ADP-ribosyl transferase (ART) subdomain which contains the active site loop (ASL) for NAD^+ binding (Figure 4-1B,D)[122,124]. Models in Figure 4-1 were derived by combining NMR models of SSB-bound Zn1 and Zn2[95], a crystal structure of the BRCT fold alone[117], and the partial full-length crystal structure of DSB-bound PARP-1[100] (PDB: 2n8a, 2cok, and 4dqy, respectively). Missing residues and linkers were filled in using MODELLER[125].

Subsequent 2012 and 2013 studies conducted by Langelier et al. set out to extrapolate the static full-length structure of PARP-1 bound to a DSB by constructing a model based on previously published atomic models of the individual PARP-1 domains[95,96,100,113-117,119,122] and additional information on key interdomain residues obtained through hydrogen-deuterium exchange mass spectrometry (HDX-MS). These studies showed that the Zn1, Zn3, WGR, and HD domains collapse around the DNA forming a continuous interface with the terminus of the DSB, leading to the activation of the CAT domain. Then in 2015, two complementary studies by Eustermann et al. and Dawicki-McKenna et al. extended this work to compose a mechanistic picture of PARP-1 during SSB repair[98,123]. These studies focused on the recognition of DNA damage by the Zn1 and Zn2 domains, deduced through nuclear magnetic resonance (NMR) and x-ray techniques, and extrapolated the subsequent collapse of the other domains based on the proposed models of DSB repair[100,101]. Notably only pieces of the PARP-1 mechanics during SSB repair are known and no full-length molecular model has been obtained directly due to its inherent flexibility causing problems for traditional techniques such as x-ray crystallography and cryogenic electron microscopy (cryo-EM).

Uninhibited, PARP-1 starts catalyzing NAD^+ to produce branching PAR chains upon binding to DNA damage, PARsylating itself (automodification) and nearby protein substrates[116,119,126] leading to the rapid recruitment of DNA repair proteins and the release of PARP-1[127]. Exploiting cellular synthetic lethality, many PARP inhibitors have been developed to compete with NAD^+ at the catalytic active site (Figure 4-1D), blocking the repair of SSBs which turn into DSBs leading to cell death in DSB repair defective tumor cells[105-107]. In 2012, Murai et al. observed that the distribution of soluble to chromatin bound PARP-1 could be shifted by PARP inhibitors calling this effect "PARP trapping"[128]. This started the debate on whether PARP inhibitor

binding leads to allosteric communication between the CAT and DNA-binding domains[89,123,124,128-130] or if DNA-binding is stabilized solely by the inhibition of automodification[120,121,131,132]. To date there have been no molecular-scale measurements of full-length PARP-1 to confirm a model of inhibitor-induced allostery.
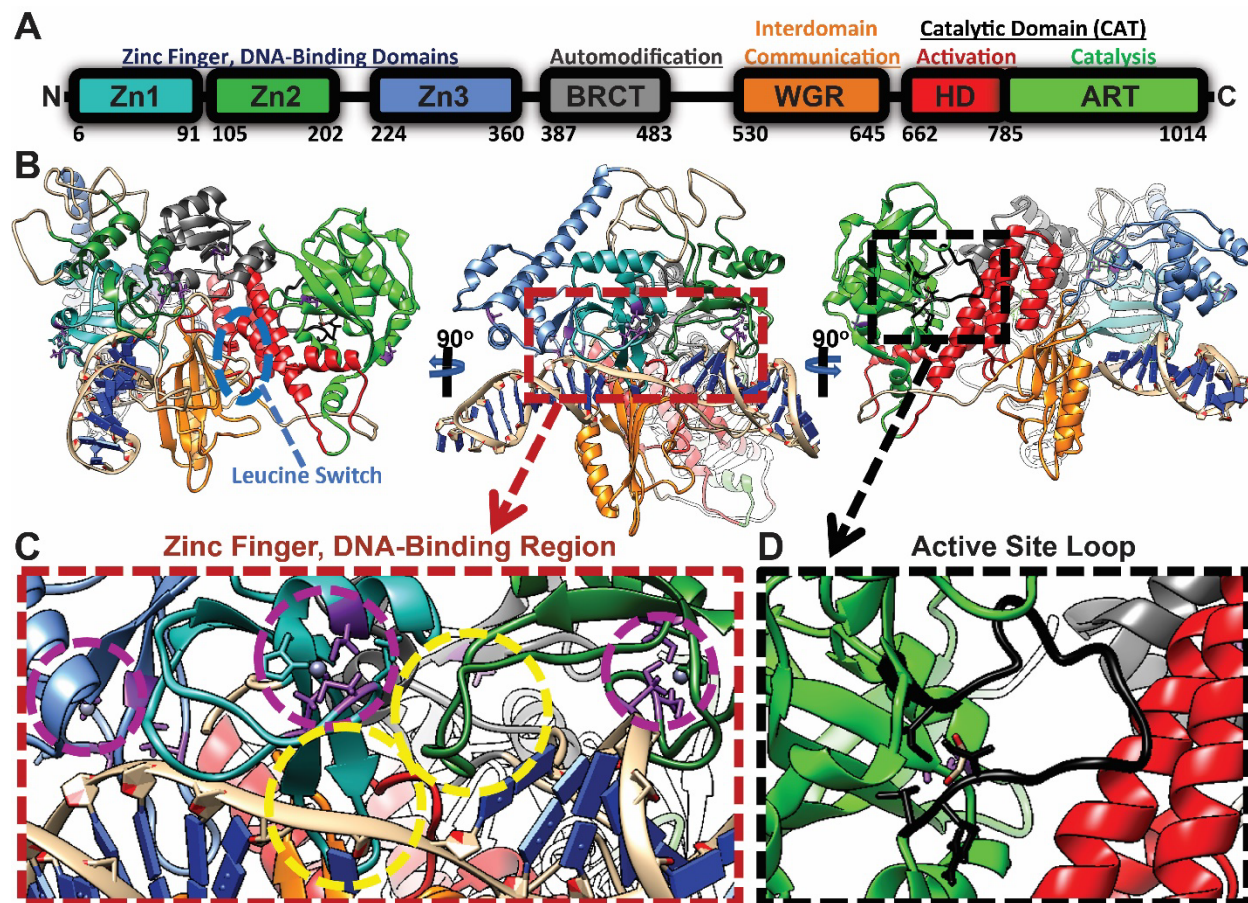
Figure 4-1. Proposed models for A) the basic architecture of PARP-1 and B) full-length PARP-1 bound to DNA during SSB repair with leucine switch (blue oval) and magnified views of the C) SSB DNA-binding region highlighting the zinc prosthetic groups (magenta) and the base stacking loops (yellow) and D) active site loop (black) in the catalytic domain.

Through XSI, dsDNA-AuNP substrates were first used as molecular rulers in the proof-of-concept works of Mathew-Fenn et al. in 2008[29,30]. These seminal works used the scattering interference pattern generated by AuNPs coupled to the ends of dsDNA to accurately and reproducibly extrapolate distance measurements of the DNA double helix. In 2013, Hura et al. applied this technique to study DNA-protein interactions by probing the mismatch repair of MutS/L via XSI[133]. In 2018, Zettl et al. were able to demonstrate the versatility of XSI by coupling AuNPs to several different classes of macromolecules and used *ab initio* mapping to determine their position in three-dimensional (3D) space[37]. Herein we adapt XSI to explore the

structure and dynamics of full-length PARP-1 molecular basis of PARP-1 damage recognition, activation, and inhibition as it is involved in SSB repair which have been previously unattainable.

4.3 Results and Discussion

4.3.1 Synthesis and characterization of DNA-AuNP substrates

Citrate-capped AuNPs were synthesized by sodium borohydride reduction followed by ligand exchange with bis(*p*-sulfonatophenyl)phenylphosphine (BSPP-AuNPs) using methods modified from previous literature (see SI Methods Section 4.6).[134,135] BSPP-AuNPs were attached to ssDNA via trithiolated linkers (Letsinger's type) on the 5'-end of ssDNA oligomers (Table 4-3), then coated with short, neutral methoxy polyethylene glycol thiol (mPEG-SH) polymers and purified, as previously described (see SI Methods Section 4.6 and Chapter 3).[12,136] Mono-functionalized ssDNA-AuNPs were isolated by anion exchange chromatography (AEX; Figure 4-5A). All AuNPs used in this study are near-spherical, with average diameters ranging from 5.8 to 7.4 nm determined from pairwise distribution functions (*P(r)*; probability plot of all inter-electron distances) obtained from SAXS for each batch of synthesized ssDNA-AuNPs (Figure 4-6A and Table 4-4). SAXS profiles for all ssDNA-AuNPs were further modeled as triaxial ellipsoidal fittings[137] using SasView software to obtain more accurate nanoparticle dimensions (Figure 4-6B, Table 4-4, and SI Methods Section 4.6). The average ratio of largest (major equatorial radius, r_A) to smallest (polar radius, r_C) particle dimension is observed at 3.4:2. Including a polydispersity parameter (defined in SI Methods Section 4.6) for r_A was essential to optimize the fit (Figure 4-6C). The average polydispersity index (PDI) of r_A is 0.18, indicating reasonable monodispersity, where PDI < 0.1 is considered ideal.[138] Complementary ssDNA-AuNPs are annealed to form dsDNA-AuNP substrates (Table 4-3) with and without SSB and purified again by AEX (see SI Methods Section 4.6 and Figure 4-7).

The final dsDNA-AuNP substrates (with and without SSB) were analyzed by XSI showing equivalent inter-AuNP. This technique is an extension of traditional solution SAXS in which a radial average of X-rays scattering off the electron density of a sample is integrated and the contrast between sample and buffer is used to produce a buffer-subtracted 1D curve in reciprocal space. The total scattering intensity is the summation of two terms: the form factor, arising from the overall particle size and morphology, and the structure factor, derived from interparticle interactions. In structural biology, SAXS samples are carefully curated to experimentally remove the contributions of the structure factor (bringing it to unity) to isolate the form factor. An inverse Fourier transform of the subtracted curves then produces *P(r)* functions, providing real-space information on the average shape of the observed electron density of individual macromolecules free from interparticle interaction. Conversely, in XSI, the structure factor is of primary importance and represents the interference pattern of scattered X-rays arising from inter-AuNP interactions, indicating discrete distances between ordered AuNPs, effectively turning the AuNPs into molecular rulers in solution (Chapter 3).[12,29,136] In an ideal XSI experiment, the scattering interference pattern between two AuNPs would be isolated by dividing out and the scattering contributions from the individual AuNPs producing a correlation scattering factor (see Extended Discussion Section 4.7).[29,30,39,82,83,136] The inverse Fourier transform of the correlation scattering factor can then be taken to give the most accurate inter-AuNP distance distribution.

Removal of the form factor from the ssDNA-AuNP-SWCNT curves did not prove beneficial and additional analyses could be performed by preserving the information from individual AuNPs. To explore AuNP size dependency, dsDNA-AuNP substrates (with and without SSB) with conjugated AuNP diameters ranging from 5.8 to 7.4 nm were analyzed by XSI (Figure 4-8). As expected, the center-to-center inter-AuNP distances followed a linear dependency proportional to the AuNP size (Figure 4-8C) with the smaller AuNP leading to reduced scattering intensity (proportional to the square of the electron density) but increased special resolution.
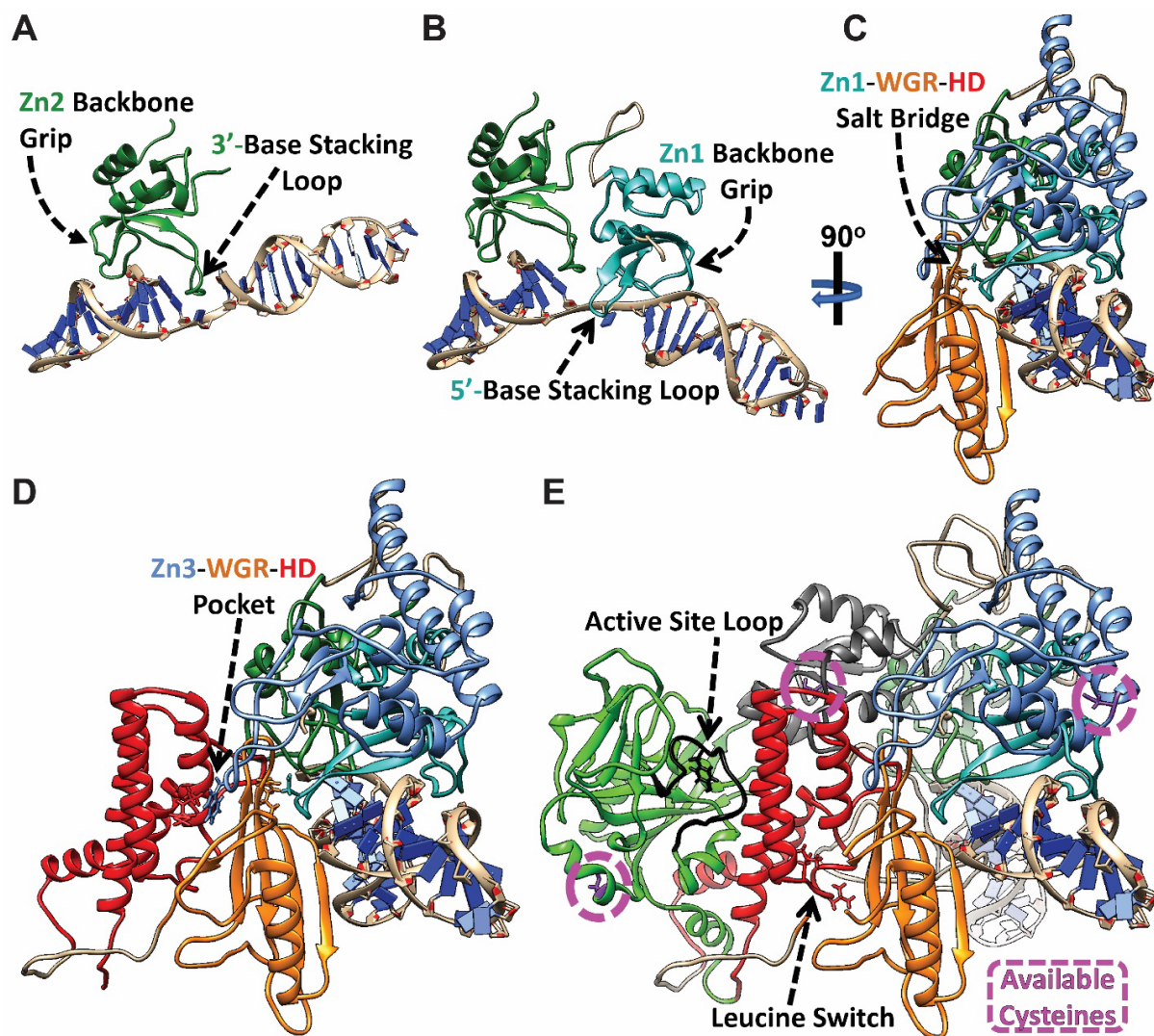
Figure 4-2. Proposed models for PARP-1 damage recognition and activation during SSB repair represented stepwise: A) Recognition of SSB by Zn2 and binding to 3'-stem, B) opening of DNA and binding of Zn1 to 5'-stem, C) Zn2 and Zn1 orient the assembly of Zn3 and WGR domains, D) WGR creates binding surface for CAT domain, and E) full assembly of PARP-1 displacing key residues in its hydrophobic core (leucine switch) leading to activation. Available cysteine residues (pink circles).

4.3.2 Molecular basis behind damage recognition and activation of PARP-1 during SSB repair using HT-XSI techniques

Previous literature has proposed models of DNA-bound PARP-1 and several mechanisms for the interaction of PARP-1 with DNA during DSB[100,101] and SSB[98,123] repair, in addition to identifying many of the critical residues associated with DNA-binding and activation[100,101]. While there are similarities in these structural models between research groups, there is substantial variation in the discussion on damage recognition and activation stemming from challenges in structural studies in solution, which have yet to be overcome. Studying the molecular basis behind the damage recognition and activation of PARP-1 during SSB repair in solution has been historically difficult due to the nature of its flexible architecture[101,112]. Thus, the structural models of DNA-bound PARP-1 and the proposed mechanics of its interactions with DNA are built on static measurements of truncated protein. Consequently, models of PARP-1 during SSB repair are principally based off those proposed through studies of DSB repair. No full-length model of PARP-1 during SSB repair has been directly measured with molecular resolution either static or time-resolved. Additionally, little is known about the structural changes to DNA during this biochemical process. Building on the previously published structural models[95,96,113-117,119,122,123] and the critical residues that have been identified[100,101], we hypothesize that the SSB-bound structure of PARP-1 will fundamentally differ from that of DSB-bound mainly due to the functional importance of the Zn2 domain in SSB repair which has been rationally excluded in previous DSB repair studies[100,101]. We build off these studies to systematically explore the molecular machinery of PARP-1 during SSB repair through truncated PARP-1 constructs measured by HT-XSI techniques.

To compose a rough, stepwise mechanism for PARP-1 damage recognition and activation during SSB repair (Figure 4-2) we have compiled information from several previous studies[98,100,101,123] and derived the models as described for Figure 4-1 (see Introduction Section 4.2) . Firstly, from two 2015 studies by Eustermann et al. and Dawicki-McKenna et al. that proposed mechanisms for PARP-1 damage recognition during SSB repair[98,123]. These works showed that the Zn1 and Zn2 domains recognize and bind SSBs in a direction specific manner, with Zn2 binding to the 3'-stem of the break (Figure 4-2A), opening the DNA and allowing Zn1 to bind to the cryptic 5'-site (Figure 4-2B). The Zn2 and Zn1 domains both use hydrophobic residues on one side of their flexible loop connecting two β-sheets to stack onto the SSB bases, and another loop that acts as a backbone grip for the minor groove of the dsDNA[100,101]. This long-distance mechanism follows a previously proposed "fly-casting" mechanism[139] that tries to explain how PARP-1 maintains binding efficiencies independently of SSB gap lengths. Currently, direct measurements of PARP-1 mechanics during SSB repair end at DNA recognition and additional information was pulled from two studies conducted by Langelier et al., which proposed a mechanism for damage recognition and activation during DSB repair[100,101]. Once bound, Zn2 and Zn1 orient the collapse of the Zn3 and WGR domains around the DNA (Figure 4-2C), which in turn stabilizes the complex through key domain interface residues (Table 4-5). Engagement of the WGR and Zn3 domains then creates a binding surface for the subsequent retraction of the CAT domain (Figure 4-2D). Notably, key residues of the WGR form a salt bridge between Zn1 and the linker between α-helices E and F on the HD subdomain producing a close connection between the DNA recognition and catalytic domains (Figure 4-2C). Similarly, the extended loop of the Zn3 zinc ribbon is engulfed by a pocket created between the WGR and the HD (Figure 4-2D), stabilizing

the helical N-terminal region of Zn3 against the turn before the C-terminal α-helix of the Zn1 domain. Importantly, these interdomain contacts alter the structure of the HD subdomain predominantly in the region around α-helix C at the WGR interface[100]. This distortion to the HD domain deforms the structure of α-helix C moving it away from α-helices B and F displacing key residues in its hydrophobic core (leucine-switch) (Figure 4-2E), suggesting that destabilization of the CAT domain is what triggers the DNA damage-induced activation of PARP-1. Binding of PARP-1 to SSBs has been demonstration to occur quickly[120] with significant association in ~10 s *in vivo*[118], while significant dissociation through automodification occurs on the order of 10 – 30 mins[131,140].
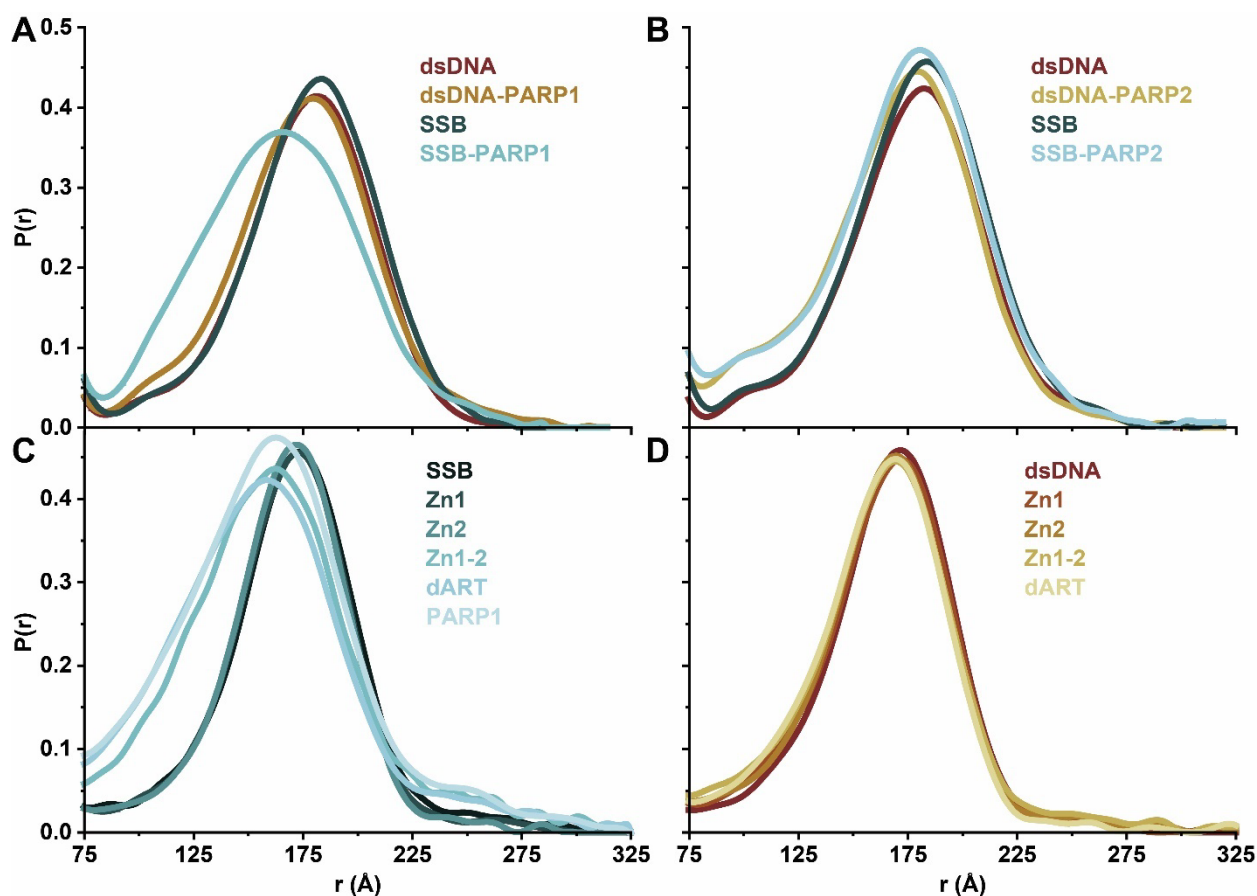
Figure 4-3. Pairwise distribution functions, *P(r)*, demonstrate an inter-AuNP distance shift only upon the interactions of PARP-1 with single-strand break (SSB) dsDNA and no movement is observed in the presence of PARP-2 or with fully duplexed dsDNA. (A-B) *P(r)* functions show inter-AuNP distance peaks for dsDNA substrates without (red series) and with a SSN (blue series) in the presence of either (A) PARP-1 and (B) PARP-2. (C-D) *P(r)* functions show inter-AuNP distance peaks for truncated versions of PARP-1 interacting with dsDNA substrates (C) with and (D) without a SSB. *P(r)* functions are normalized to the primary intra-AuNP peak, then this peak is omitted for clarity (not on the x-scale). Corresponding scattering profiles are shown in Figure 4-9.

Protein interactions with the dsDNA-AuNP substrates were measured *via* XSI in solution (Figure 4-3, Figure 4-9, and SI Methods Section 4.6). Real-space analysis of the scattering profiles produces *P(r)* functions with two main peaks. The first peak represents the intra-AuNP distances between electrons within individual AuNPs, with the peak maximum being the average radius of the AuNPs. The sole presence of this peak signifies the absence of long-range order (i.e., free AuNPs in solution). The intra-AuNP peak provides a reference for the AuNP size distribution in each sample and enables normalization between samples to account for slight fluctuations in concentration and X-ray beam intensity. The second broader peak in the *P(r)* functions represents the inter-AuNP distances which only appears for AuNPs held at discrete distances (i.e., dsDNA-AuNPs not ssDNA-AuNPs) and shifts upon protein interaction with decrease in inter-AuNP distances representing bending of the dsDNA (Figure 4-3).

As expected, only SSB dsDNA-AuNPs bent upon interaction with PARP-1 (Figure 4-3A) and no interaction was observed for PARP-2 (Figure 4-3B). Additionally, several truncated versions of PARP-1 were examined showing sequential bending as domains are added (Figure 4-3C and Table 4-1). Importantly, negligible bending was observed for dsDNA-AuNP absent SSB (Figure 4-3D and Table 4-1). It is predicted that Zn2 binds (Figure 4-2A), opening the DNA and allowing Zn1 to bind (Figure 4-2B). Experimentally we observe the decrease inter-AuNP distances for Zn1 and Zn2 alone to be negligible for both dsDNA- and SSB-AuNPs. However, upon addition of Zn1 and 2 (Zn1-2) together we observe a drastic decrease in the inter-AuNP distance of 9.9 ± 2.2 Å only in the presence of a SSB demonstrating that both domains are needed for the initial opening of the SSB. After Zn1-2 binding, the WGR and Zn3 domains is then expected to create a binding surface for the CAT domain (Figure 4-2D). This construct missing the ART domain (dART) showed further bending of the dsDNA with a decrease in inter-AuNP distance of 13.3 ± 1.3 Å only in the presence of a SSB. Finally, when the complete CAT domain is added back to full length PARP-1, it showed a slight increase in the inter-AuNP distance for SSB dsDNA-AuNPs back to 11.4 ± 3.7 Å suggesting a slight relaxation of the dsDNA. Importantly however, the inter-AuNP peak for PARP-1 with a SSB demonstrates a broadening to encompass all related peaks suggesting that the slight increase in the inter-AuNP distance reflects an increased dissociation constant for full length PARP-1 with an active CAT domain (Figure 4-3C).

Table 4-1. Decrease in inter-AuNP distances in the presence of protein obtained from pairwise distribution functions of X-ray scattering.

	Inter-AuNP Distance Shift (Å)	
	dsDNA	SSB
Zn1	1.8 ± 1.3	0.0 ± 0.5
Zn2	3.8 ± 0.9	1.8 ± 0.8
Zn1-2	3.7 ± 0.5	9.9 ± 2.2
dART	4.2 ± 1.0	13.3 ± 1.3
PARP1	4.7 ± 1.9	11.4 ± 3.7
PARP2	3.1 ± 0.3	3.1 ± 0.3

4.3.3 *Define the allosteric basis behind PARP-1 inhibition during SSB repair using HT-XSI techniques*

Although PARP inhibitors are all small molecules that competitively bind to the catalytic active site, they exhibit vastly different effects on PARP-1 affinity for DNA damage, thus creating perhaps the most significant paradox in the study of PARP-1 drug design[130]. We explore the presence of allosteric changes associated with PARP inhibitor binding during SSB repair through HT-XSI.

Despite a deficiency in direct measurements researchers have theorized that variations in PARP inhibitor binding at the catalytic active site influences PARP-1 affinity for DNA through allosteric changes[128] while others argue that this dependency is solely due to catalytic inhibition[120]. Based on previously published mechanistic studies[89,123,124,128,130], we hypothesize that the observed effects on PARP-1 retention are derived from allosteric communication between the catalytic and DNA-binding domains triggered by slight differences in the molecular structure of the PARP inhibitors and their unique interactions with the catalytic active site, which will also exhibit measurable structural changes to the DNA. We believe that these changes can be defined through HT-XSI measurements using AuNPs conjugated to SSB dsDNA-AuNP substrates or directly to PARP-1 made in the presence of inhibitors, in conjunction with the systematic site-directed mutagenesis of critical residues between domains and within the active site.

Promising models of PARP-1 inhibition during DSB repair suggest that allosteric communication between the CAT domain and the DNA binding domains may be what leads to variation in affinity and retention of PARP-1 at the break site[124]. These PARP inhibitors have been broken down into three categories[89]: (i) Type 1 inhibitors produce a strong allosteric change, destabilizing the HD and leading to slower release, (ii) Type 2 inhibitors lead to little to no allosteric or affinity changes, and (iii) Type 3 inhibitors produce a strong allosteric change that conversely stabilizes the HD leading to faster release. In 2012, Murai et al. were the first group to show that the distribution of soluble to chromatin bound PARP-1 could be shifted by PARP inhibitors calling this effect "PARP trapping"[128]. This study theorized that there must be variations in the binding modes of the different PARP inhibitors leading to a "reverse allostery" influencing PARP retention on DNA. Recent studies have used molecular dynamics (MD) to simulate the allosteric changes caused by five of the most active clinical inhibitors[129,130], however, there are still very few direct measurements of this allosteric communication. Two key studies have lent support to this allosteric model by: (i) exploring how PARP-1 allosteric activation can be manipulated by occupying the NAD^+ binding site with non-hydrolyzable NAD^+ analogs[124] and (ii) showing that PARP-1 without the HD retains the same catalytic activity as if it were bound to a DNA break, demonstrating the autoinhibitory effects of the HD[123]. Conversely several studies have used fluorescence-based techniques to negate the theory of allostery, postulating that DNA-binding is stabilized solely by the inhibition of automodification[120,121,131,132].
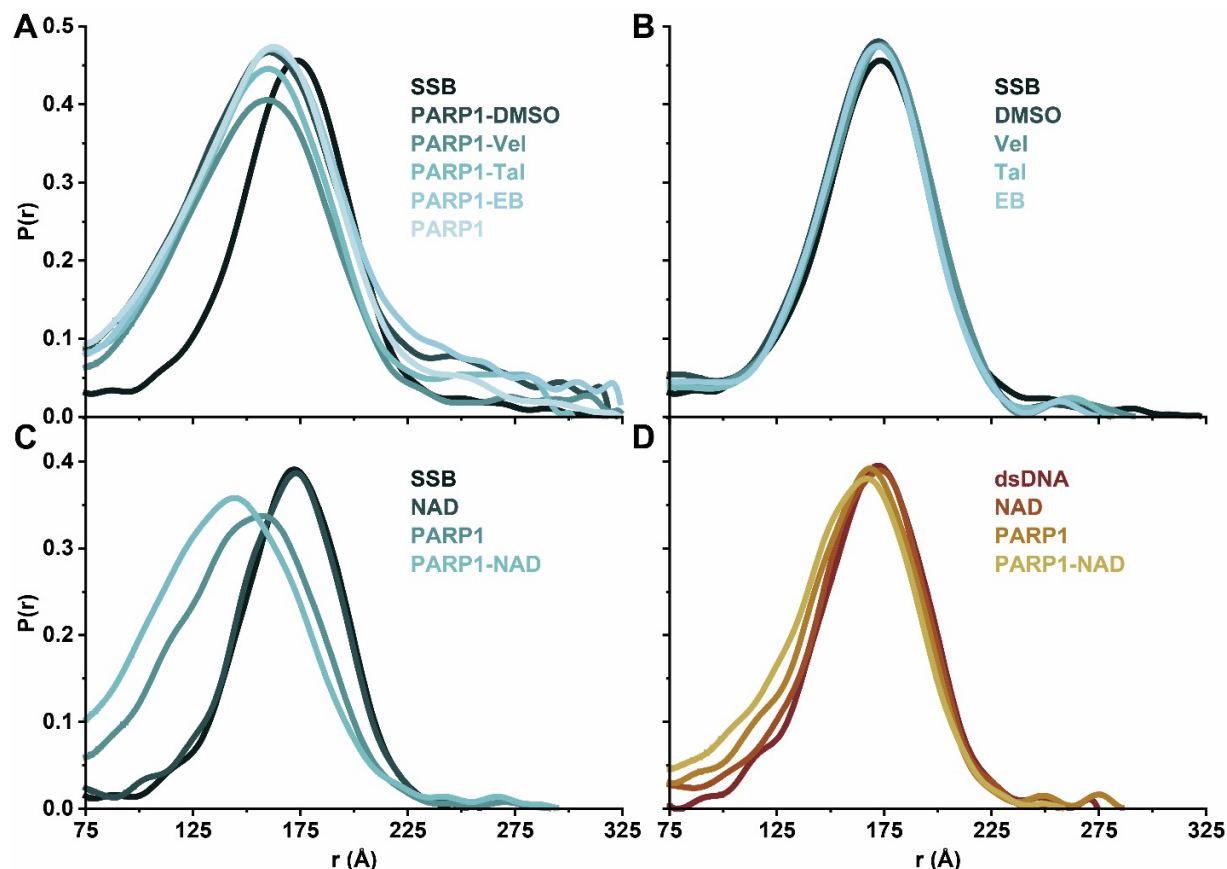
Figure 4-4. Pairwise distribution functions, *P(r)*, demonstrate that the inter-AuNP distance shift of PARP-1 interaction with SSB dsDNA remains unchanged by the presence of known PARP-1 inhibitors but shows a notable shift in the presence of NAD^+. *P(r)* functions show inter-AuNP distance peaks for (A-B) single-strand break (SSB) dsDNA substrates in the presence of known PARP-1 inhibitors Veliparib (Vel), Talazoparib (Tal), and EB-47 (EB), (A) with and (B) without PARP-1 and (C-D) PARP-1 interaction with (C) SSB dsDNA and (D) dsDNA in the presence of NAD^+. *P(r)* functions are normalized to the primary intra-AuNP peak, then this peak is omitted for clarity (not on the x-scale). Corresponding scattering profiles are shown in Figure 4-10.

PARP-1 interactions with the dsDNA-AuNP substrates were measured *via* XSI in solution in the presence of several known PARP inhibitors as well as NAD^+ (Figure 4-4, Figure 4-10, and Table 4-2). Three PARP inhibitors were tried against SSB- and dsDNA-AuNPs: Veliparib (Vel), Talazoparib (Tal), and EB-47 (EB) (Figure 4-4A and Table 4-2). Dimethylsulfoxide (DMSO) was also studied as a control as all three inhibitors required 1% DMSO to stabilize them in aqueous solution. As expected, all three inhibitors (and DMSO) showed negligible change in inter-AuNP distances on dsDNA-AuNPs, or either substrate without PARP-1 (Figure 4-4B). All three inhibitors (and DMSO) in the presence of PARP-1 and SSB dsDNA-AuNPs showed nearly the same decrease in inter-AuNPs as PARP-1 without inhibitors to average 12.6 ± 3.5 Å. Interestingly, only Vel and Tal showed a notable narrowing of the inter-AuNP peak (with Vel showing the greatest narrowing) suggesting that the inhibitors may be lowering the dissociation

constant of PARP-1 bound to the SSB. PARP-1 interaction was also studied with both dsDNA-AuNP substrates in the presence of NAD^+, which is hypothesized to increase the dissociation of PARP-1 through automodification, relaxing the dsDNA, broadening the inter-AuNP peak and leading to increased inter-AuNP distances. Surprisingly, PARP-1 bound to SSB dsDNA-AuNP with NAD^+ did show some inter-AuNP peak broadening but the inter-AuNP distances decreased further than PARP-1 alone to 27.6 ± 1.6 Å suggesting increased bending of the dsDNA (Figure 4-4C and Table 4-2). No such changes are observed with NAD^+ for dsDNA-AuNPs, or either substrate without PARP-1 (Figure 4-4C-D).

Table 4-2. Decrease in inter-AuNP distances in the presence of PARP-1 upon the addition of NAD^+ or Inhibitors obtained from pairwise distribution functions of X-ray scattering.

	Inter-AuNP Distance Shift (Å)	
	dsDNA	SSB
PARP1 (alone)	4.7 ± 1.9	11.4 ± 3.7
Veliparib (Vel)	2.0 ± 1.3	14.1 ± 1.4
Talazoparib (Tal)	4.8 ± 0.6	15.4 ± 2.1
EB-47 (EB)	4.4 ± 1.2	11.8 ± 2.8
DMSO (control)	6.1 ± 0.0	14.4 ± 2.4
NAD⁺	4.8 ± 3.6	27.6 ± 1.6

4.4   Conclusion

In this work, we demonstrate that HT-XSI is a valuable technique for studying DNA-protein interactions in solution, using specially designed dsDNA-AuNP substrates to act as molecular rulers. XSI harnesses the electron-rich gold atoms in AuNPs to enable the study of these systems at greatly reduced concentrations saving considerable material which is often extremely difficult to prepare. Herein, we design dsDNA-AuNP substrates with and without a SSB. We analyze the DNA-protein interactions between these specially designed substrates and several truncated and full-length constructs of PARP-1. These construct studies provide insight into the molecular basis behind damage recognition and activation of PARP-1. The first notable interaction (binding/bending) of the SSB dsDNA-AuNPs occurs upon addition of Zn1 and 2 together with a drastic decrease in the inter-AuNP distance of 9.9 ± 2.2 Å. Then upon addition of the WGR, Zn3, and HD domains (dART construct) a further decrease in the inter-AuNP distances is observed to 13.3 ± 1.3 Å. Final with full length PARP-1, a slight increase in the inter-AuNP distance for SSB dsDNA-AuNPs back to 11.4 ± 3.7 Å and a peak broadening is observed suggesting a slight relaxation of the dsDNA and an increased dissociation constant for full length PARP-1.

Full length PARP-1 interaction with SSB dsDNA-AuNPs was also measured in the presence of three known PARP-1 inhibitors (Vel, Tal, and EB) as well as NAD^+ necessary for automodification of the CAT domain. All three inhibitors (and DMSO) in the presence of PARP-1 and SSB dsDNA-AuNPs showed nearly the same decrease in inter-AuNPs as PARP-1 without

inhibitors to average 12.6 ± 3.5 Å with Vel and Tal showed a notable narrowing of the inter-AuNP peak (with Vel showing the greatest narrowing) suggesting that the inhibitors may be lowering the dissociation constant of PARP-1 bound to the SSB. Surprisingly, PARP-1 bound to SSB dsDNA-AuNP with NAD^+ showed inter-AuNP distances decreasing further than PARP-1 alone to 27.6 ± 1.6 Å suggesting greatly increased bending of the dsDNA.

These results are as yet preliminary and much has yet to be explored about the molecular basis of PARP-1 damage recognition, activation, and inhibition during SSB repair using this technique (see Future Directions Section 4.5 and Extended Discussion Section 4.7). As such HT-XSI has proven to be a viable methodology to studying these systems and indeed most DNA-protein interactions.

4.5   Future Directions

4.5.1   HT-XSI assay for PARP-1 SSB-binding time-resolved dynamics studies

Time-resolved studies will be conducted by running a series of pump-probe experiments to determine the mechanism of dissociation through automodification since this process is known to occur over far longer timescales than association[118,131,140]. These experiments will involve the *in situ* activation of PARP-1 bound to the SSB, through the photo-activation of synthesized 1-(4,5-dimethoxy-2-nitrophenyl)ethyl (DMNPE) or 1-(2-nitrophenyl)ethyl (NPE)-caged NAD^+ molecules[141,142] by a secondary, tunable UV-vis light source orthogonal to the primary x-ray beam.

4.5.2   HT-XSI assay of site-directed PARP-1 mutants

Researchers have systematically identified many critical residues at the interface with DNA[98,100,123], between domains[100,101], and within the active site[89,123], many of which make intriguing mutagenic targets for investigating the molecular basis behind DNA-binding, interdomain communication, and activation. We have tabulated many of these proposed residues of interest (Table 4-5) to use as a map to systematically explore the static structure and time-resolved dynamics of PARP-1 SSB-binding and activation through site-directed mutagenesis in conjunction with HT-XSI.

4.5.3   HT-XSI assays of inhibitor influence on PARP-1 allosteric communication and DNA structural changes during SSB repair

We believe that the subtle structural changes of the DNA, during binding and subsequent release of PARP-1, can be visualized and that these changes will be indicative of the allosteric changes associated with variations in the binding modes of the different PARP inhibitors at the catalytic active site. PARP-1 bound to SSB dsDNA-AuNP substrates will be screened in the presence of a large "promiscuous" inhibitor library. Hits from the full-length PARP-1 assay will then be run against truncated constructs bound to SSB dsDNA-AuNP substrates. Additionally, the labeling of strategic cysteine residues on PARP-1 with 0.7 nm AuNPs[37], will certainly be required to obtain the more minor allosteric changes suggested by MD simulations[129,130]. PARP-1 has one

accessible cysteine residue on both the ART and Zn3 domains (Figure 4-1C) and each would need to be labeled to properly visualize allosteric communication between the catalytic domain and the DNA-binding domains.

4.5.4   HT-XSI assay of inhibitor influence on site-directed PARP-1 mutants

Researcher have systematically identified many potential residues which may be involved in the allosteric changes during PARP-1 inhibition[89,123,130,143,144]. Since a full-length crystal structure of DNA-bound PARP-1 in the presence of inhibitors is currently unavailable, due to difficulties in co-crystallization, a recent study by Kumar et al. has used MD to simulate the allosteric changes caused by five of the most active clinical inhibitors[130]. This study was performed by docking these small molecule PARP inhibitors to a derived partial full-length structure, without the Zn2 and BRCT domains, and was able to identify many residues potentially rearranged by the binding of inhibitors within the catalytic active site or involved in allosteric communication between domains. We have tabulated many of these proposed residues of interest (Table 4-6) to use as a map to systematically explore PARP-1 inhibitor induced allosteric changes through site-directed mutagenesis and HT-XSI.

4.6   Materials and Methods

Materials

5'-trithiolated-ssDNA (Letsinger's type) oligonucleotides with SDS-PAGE purification were purchased from Fidelity Systems (Gaithersburg, MD). Methoxy poly(ethylene glycol) thiol (mPEG-SH; MW ~350 g/mol) was purchased from Biochempeg Scientific Inc. (Watertown, MA). All other reagents were purchased from Millipore Sigma (St. Louis, MO).

Table 4-3. Table showing DNA substrate sequences.

31-bp-Fully Duplexed dsDNA	5'-(AuNP)-TrT-GCC TTC ATA TCT GGA GTG CTT ATT CCT ATC G-3' 3'-CGG AAG TAT AGA CCT CAC GAA TAA GGA TAG C-TrT-(AuNP)-5'
31-bp-Single-Strand Break SSB dsDNA	5'-(AuNP)-TrT-GCC TTC ATA TCT GGA GTG CTT ATT CCT ATC G-3' 3'-CGG AAG TAT AGA CCT C*AC GAA TAA GGA TAG C-TrT-(AuNP)-5'
TrT = Trithiol linker (Letsinger's type), *missing phosphate group	

Methods

Synthesis of Citrate-Capped Gold Nanoparticles (AuNPs)

Citrate-capped AuNPs of diameters 5.8-7.4 nm were prepared using a method modified from that which was previously described.[134] Briefly, a 2 L solution of 0.25 mM $HAuCl_4$ and 0.25 mM tri-sodium citrate was prepared in a conical flask using ddH_2O cooled to 4°C. Next, 10 mL of 0.6 M

NaBH$_4$ at 4°C was added rapidly to the solution while stirring. The solution turned dark red immediately after adding NaBH$_4$, indicating particle formation. The solution was allowed to warm to room temperature (RT) and stirred overnight for the water to decompose excess NaBH$_4$.

Citrate-BSPP Exchange for Gold Nanoparticles (BSPP-AuNPs)

Bis-(*p*-sulfonatophenyl) phenylphosphine (BSPP) was added to citrate-stabilized colloidal AuNPs (~5.0 x 10^{13} particles/mL) to a final concentration of 0.5 g/L and stirred at RT for a minimum of 6 hours. Approximately 1 mL of saturated NaCl solution was added per 10 mL BSPP-exchanged colloidal gold, until the solution changed from transparent red to a darker, cloudy purple, indicating the reversible precipitation of the AuNPs.[135] The mixture was centrifuged (Beckman, JA-18 rotor) at 12,000 rcf for 10 min and decanted carefully to not disturb the pelleted BSPP-AuNPs. BSPP-AuNPs were washed twice with 0.5 M NaCl solution (repeating the centrifugation step above) and resuspended in 15 mM phosphate buffer, 1 mM TCEP, pH 7 for storage. Suspensions were stored at 4°C until use. Note that freezing caused sample precipitation.

Conjugation of Single-Stranded DNA to BSPP-AuNPs (ssDNA-AuNPs)

If stored longer than 2 weeks, fresh TCEP was added to reduce the solution of colloidal BSPP-AuNPs prior to ssDNA conjugation. To do this, saturated NaCl solution was first added to BSPP-AuNPs until the solution turned dark, then the BSPP-AuNPs were centrifuged at 12,000 rcf for 10 min and resuspended in 15 mM phosphate buffer, 1 mM TCEP, pH 7. The final BSPP-AuNP concentration was determined by measuring the absorbance at 520 nm (NanoDrop 2000, Thermo Scientific) and converting to concentration with the empirical extinction coefficient,[145] ε_{520nm} = 9.69x10^6 L mol^{-1} cm^{-1}. The concentration of desired 5'-trithiolated-ssDNA (Letsinger's type) oligonucleotides was calculated by measuring the absorbance at 260 nm using the sequence-dependent extinction coefficient. Solutions of BSPP-AuNPs and trithiolated-ssDNA were mixed vigorously at a final molar ratio of 1:1 and incubated at RT overnight. A 129 mM mPEG-SH solution was prepared and added to the ssDNA-AuNP suspension at a final molar ratio of 3000:1 mPEG-SH to AuNPs. The final ssDNA-AuNP concentration was determined again by measuring absorbance.

AuNPs were coated with mPEG-SH to prevent aggregation and aid in the purification by anion exchange chromatography.[136] After mPEG-SH coating, the hydrodynamic radii of the AuNPs measured by DLS increase from the AuNP core radii measured with SAXS by an average of 1.76 ± 0.13 nm across all samples (Figure 4-6D and Table 4-7). This increase in hydrodynamic diameter implies successful mPEG-SH functionalization of the exposed AuNP surfaces.[146] The average length of the mPEG-SH (MW ~350) is ~1.7 nm when fully extended, based on the length of repeating polyethylene oxide units (0.278 nm, n=6) in water.[147] However, even if all available mPEG-SH functionalized the surface, the theoretical distance between grafting sites (~5.2 nm) would far exceed the calculated Flory Radius (0.815 nm), leading to only a partially extended mushroom conformation. This result is in line with the measured change in hydrodynamic radius imparted by mPEG-SH being slightly less than the expected change based on polymer length.[148-150]

Anion Exchange Chromatography (AEX) Purification of ssDNA-AuNPs

Mono-conjugated ssDNA-AuNPs were isolated as previously described[136] using a Dionex DNA-Pac PA100 anion exchange column on either a GE AKTA Explorer or a GE Atka Pure fast protein liquid chromatography (FPLC) with an NaCl gradient from 0.01 to 1 M over a period of 55 min at a flow rate of 1.1 mL/min. Sample elution was monitored by measuring UV-Vis absorption at 260 (ssDNA) and 520 nm (AuNPs), and the mono-conjugated ssDNA-AuNP fraction was collected for downstream use (Figure 4-5).

dsDNA-AuNP Annealing, Secondary AEX Purification, Dialysis

Complementary ssDNA-AuNP conjugates are annealed by mixing complementary AEX purified mono-conjugated ssDNA-AuNPs at equal molar ratios in 1.5 mL Eppendorf tube. Mixture is then heated at 94 C for 3 min and allowing to cool to RT slowly to form dsDNA or SSB dsDNA substrates as shown in Table 4-3. The duplexed dsDNA then undergoes secondary AEX purification to isolate only annealed dsDNA as shown in Figure 4-5C-D. Final dsDNA-AuNP substrates were dialyzed overnight in 20 mM Hepes pH 8, 150 mM NaCl to enable PARP-1 interaction. Importantly, PARP-1 SSB interaction was found to be salt dependent hitting an affinity peak at 150 mM salt (Figure 4-11).

Characterization by Dynamic Light Scattering (DLS)

DLS measurements were taken with the Zetasizer Nano ZS (Malvern Analytical) with a material refractive index of 0.200 and absorption of 3.320 for colloidal gold.[151,152] All samples were diluted in 0.1X PBS to an AuNP concentration of 0.20-0.25 µM and loaded in disposable cuvettes (Malvern ZEN0040) for size measurement.

Structure Studies by High-Throughput X-ray Scattering Interferometry (HT-XSI)

HT-XSI data was collected at the SIBYLS beamline (bl12.3.1) at the Advanced Light Source of Lawrence Berkeley National Laboratory, Berkeley, California.[2] X-ray wavelength was set at $\lambda =$ 0.12398 nm and the sample-to-detector distance was 2.07 m, resulting in a scattering vector (q) range of 0.1 - 4.6 nm^{-1}, which corresponds to real-space distances of 62.8 - 1.4 nm. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, with scattering angle 2θ. Data was collected using a Dectris PILATUS3X 2M detector at 20°C and processed as described previously.[10]

Immediately prior to data collection, 15 µL of each sample was added to 15 µL of buffer in a 96-well plate kept at 10°C for final corresponding concentrations of 250 nM dsDNA-AuNPs and 500 nM protein. If inhibitors or nicotinamide adenine dinucleotide (NAD$^+$) are used, they are added at a final concentration of 5 mM. Additionally, up to 1% dimethylsulfoxide (DMSO) v/v may be required to stabilize inhibitors in aqueous solution. Each sample was then transferred to the XSI sampling position via a Tecan Evo liquid handling robot (Tecan Trading AG, Switzerland) with modified pipetting needles acting as sample cells as described previously.[43] Samples were exposed to X-ray synchrotron radiation for 5 s at a 0.1 s frame rate for a total of 50 images. Each collected image was circularly integrated and normalized for beam intensity to

generate a one-dimensional scattering profile. Buffer subtraction was performed for the one-dimensional scattering profile of each sample using each of two bracketing buffer wells to ensure the subtraction process was not subject to instrument variations. Scattering profiles over the 5 s exposure were sequentially averaged together to eliminate any potential radiation damage effects. All data processing was done using our beamline specific data processing pipeline by the SIBYLS SAXS Process (SSP) GUI (see Sections 2.6 and 3.4).[12,42] Pairwise distribution functions, *P(r)*, were generated in batch using the automated GNOM[44] feature of the SSP GUI.

Triaxial Ellipsoidal Fitting

Triaxial ellipsoidal fittings of AuNP scattering curves were performed using SasView by incorporating the classical physical properties of the light scattering of ellipsoidal particles in solution.[137] Here, the scattering of randomly ordered particles in solution is expressed as follows:

$$I(q) = scale(\Delta\rho)^2 \frac{V_{NP}}{4\pi} \int_\Omega \phi^2(qr)\, d\Omega + background$$

4-1

Where the integral sums all possible rotational orientations in solution, $\Delta\rho$ is the scattering length density difference between the sample and buffer and V_{NP} is defined as the volume of an AuNP (V_{NP}) as follows:

$$V_{NP} = \frac{4\pi}{3} r_A r_B r_C$$

4-2

For the triaxial ellipsoidal fittings in this work, the buffer was 0.1X PBS, with a scattering length density close to ρ_{water} at room temperature (9.44 x 10^{-6} Å^{-2}) used as the starting value of the fitting. Other fit parameters were left unconstrained and the DREAM algorithm[153] was used to fit the data. The fitting was further optimized by allowing for the polydispersity of the major equatorial radius (r_A) to be modeled as a Gaussian distribution, where the polydispersity index, PDI = standard deviation / mean (Figure 4-6C). Fittings are shown in Figure 4-6B and data are tabulated in Table 4-4.

4.7   Extended Discussion

4.7.1   *Early Data with the Isolation of the Correlation Scattering Factor (CSF)*

After data collection and initial data process (see Methods Section 4.6) , Mathematica[154] or MatLab[155] programs, depending on the size of the AuNPs, can be used to process the averaged SAXS curves in order to obtain accurate distance measurements. In the case of 5 nm AuNPs the SAXS profiles can be directly analyzed using the Debye approximation for systems of equivalent particles as previously described[80]. First the scattering intensity of each SAXS curve in the experimental series and the homogeneous solution of the AuNPs used to create the dsDNA-AuNPs substrates are scaled over the range q = 0.1 to 0.2 Å^{-1} where the inter-gold correlation

scattering is minimal. Each scaled SAXS profile is divided by a theoretical scattering curve generated by fitting the colloidal gold NP profile to a triaxial ellipsoid model. This eliminates the intra-gold scattering terms and outputs a correlation scattering factor (CSF). Alternatively, to generate a CSF for samples with AuNPs of diameter <4 nm, the inter-gold scattering intensity must be first isolated from the cross-scattering terms between the gold and any macromolecules in solution in addition to the intra-gold contribution by collecting scattering controls from all the appropriate standards and calculating the scaling factors for each one as previously described[39]. The CSF is a damped sine wave function that oscillates across the median at $q<0.1$ Å^{-1} and must be manipulated to oscillate around zero by subtracting a constant parameter (k_2) over the entire q range, which is determined from fitting the local maxima as $1/(k_1q) + k_2$. Each CSF is then transformed via the GIFT algorithm[33], which fits the CSF as a set of cubic B-spline basis functions to produce a distribution function of inter-gold distances, $P(D_{i,j})$. In the case of using smaller AuNPs, coupled directly to the DNA-binding protein, as secondary reference scatters, several additional analysis steps must be taken to map the NP positions in 3D space, as previously described[37].

We have been able to reproducibly prepare and collect scattering profiles for 5 nm gold coupled 37-mer dsDNA-AuNP substrates with and without a SSB (Figure 4-12A). Automated processing was done and initial distribution functions, P(r), were generated straight from averaged scattering profiles for each dsDNA-AuNP substrate alone and in the presence of PARP-1 (Figure 4-12B). Then the data was run again through the whole processing and analysis pipeline to produce CSFs (Figure 4-12C) and distribution functions, $P(D_{i,j})$, eliminating the intra-gold scattering contributions. Results demonstrated a clear difference in the inter-gold distances only in the case of PARP-1 interaction with the SSB dsDNA-AuNP substrate, confirming its effectiveness for this application. The observed shift in the peak of the distribution functions, $P(D_{i,j})$ and the shift to higher q in the CSFs represent a decrease in the inter-gold distances suggesting bending of the dsDNA upon PARP-1 binding to the SSB.

Distribution functions, $P(D_{i,j})$, and CSFs for each have been generated as represented by Figure 4-12D and E respectively. These plots show fluctuations in the inter-gold distances between samples of PARP-1 bound to SSB dsDNA-AuNP substrates in the presence of inhibitors. Out of the five PARP inhibitors tested, three of them showed little to no effect on SSB-bound PARP-1 as compared to the SSB-bound PARP-1 with DMSO used as a control. Conversely, UKTT15 (a variant of veliparib) showed a slight shift to lower q in the CSF suggesting a straightening of the DNA bend and EB-47 showed an extension of the inter-gold distance beyond that of the original SSB dsDNA-AuNP substrate (Figure 4-12D, E). UKTT15, which is known to be a strong type I inhibitor, should greatly increase the binding affinity of PARP-1 for DNA, but these results suggest a straightening of the DNA-bend and a weakening of the PARP-1 binding affinity. As for EB-47, which is thought to closely mimic the binding of NAD$^+$ within the CAT active site[89], may be promoting the binding of multiple PARP-1 monomers to a single dsDNA leading to a slight uncoiling of the helix causing the observed extension in the DNA.

4.7.2   Benchmark, Potential Problems, and Alternative Approaches

By completing preliminary and conditional screenings we will establish initial experimental parameters and optimize the HT-XSI pipeline to suit static and time-resolved SSB repair studies

on native PARP-1 and its site-directed mutants. We may find that measuring the DNA dynamics is insufficient to interpret the details of some subtle DNA-protein interactions, thus requiring us to use 0.7 nm AuNPs coupled directly to available cysteine residues on PARP-1 (Figure 4-2E) as secondary reference scatters to obtain orientation information, aiding in 3D reconstruction of the system. The major challenge with this method is that ideally only one or at maximum two cysteine residues are accessible for coupling without perturbing the system, otherwise orientational information becomes convoluted. There are seven cysteine residues on PARP-1. Three covalently bound to zinc (Figure 4-1C) and a fourth in the ASL which are all inaccessible, one on the BRCT domain, which can be modified without disrupting PARP-1 functionality[100,113], and one on both the Zn3 and ART domains. Depending on the goal of the study, these residues should provide a sufficient means to provide secondary reference scattering. A concern for the proposed time-resolved studies is the damage to the samples caused by prolonged exposure to synchrotron radiation. To compensate for radiation damage, a continuous exchange of material may be required, which can be accomplish by adapting the custom flow cell designed for the SEC-SAXS-MALS system described previously (see Section 2.3.2)[3-5,7,8,156]. Additionally, our proposed time-resolved studies involve the measurement of PARP-1 dissociation from the dsDNA-AuNP substrate which cannot be assumed to translate directly to its binding to DNA damage. Thus, to measure the much faster time dynamics of PARP-1 SSB-binding the SEC-SAXS-MALS flow cell can be modified to do simple mixing experiments but any use of this flow cell would require significantly more material.

The majority of studied PARP inhibitor compounds need to be dissolved in DMSO in order to stabilize them in solution but can lead to increased sensitivity of the samples to x-ray radiation even at concentrations as low as 1% v/v. The concentrations used in our preliminary results already suggest strongly increased susceptibility to radiation damage causing aggregation of the system after only three frames (0.6 s) in addition to slight relaxation of the DNA-bend in the system we are studying (Figure 4-12D, E). While this is fine for static inquiry, this issue must be overcome before proceeding to longer exposures. To compensate for radiation damage, the continuous exchange of material may be require which can be accomplish by adapting the custom flow cell designed for the SEC-SAXS-MALS system described previously (see Section 2.3.2)[3-5,7,8,156]. Alternatively, the x-ray flux may be attenuated, or the photon energy increased to reduce sample interaction. As the allosteric changes to PARP-1 may be far too small to properly detect through low resolution x-ray scattering techniques additional measurements may be required. For SSB-bound PARP-1 systems that demonstrate consistent structural changes in the presence of inhibitors, cryo-EM may be used to obtain higher resolution electron density maps which can be combined with SEC-SAXS-MALS measurements to model dynamic changes in solution[4].

4.8 Chapter Supporting Information

4.8.1 Supplementary Figures and Tables
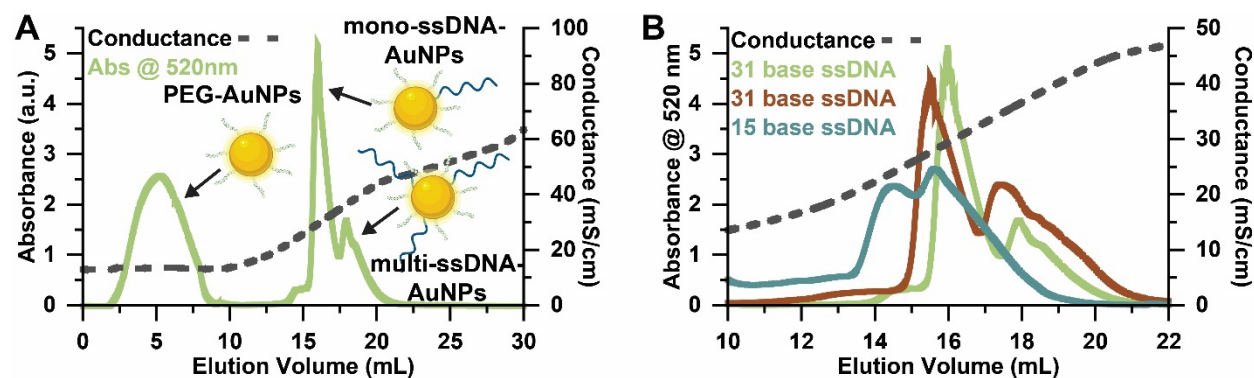
Figure 4-5. Anion exchange chromatograms for (A) the separation of non-conjugated PEG-AuNPs, mono-conjugated ssDNA-AuNP, and multi-conjugated ssDNA-AuNP using fast protein liquid chromatography (FPLC) and (B) separation of ssDNA-AuNPs used to make dsDNA substrates, demonstrating the length-dependent shift in the elution volume. Spectra are measured by diode array detector (DAD) at 520 nm. Conductance measurements depict salt gradient conditions (dash grey). Fractionation shown by enumerated vertical lines.
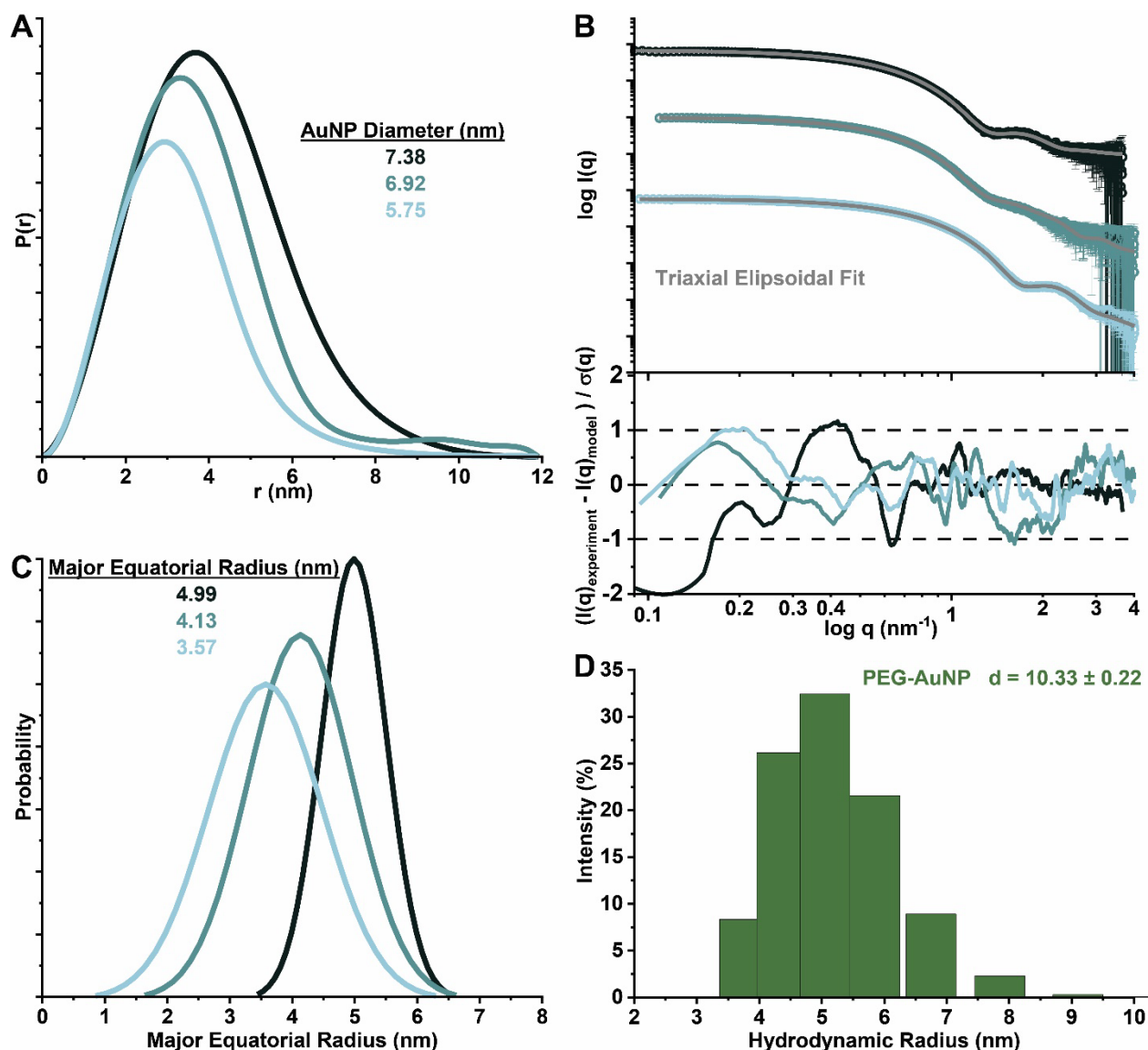
Figure 4-6. SAXS analysis for synthesized ssDNA-AuNPs. (A) Calculated *P(r)* functions for AuNPs of various sizes, scaled to the calculated average diameters for visual clarity. (B) Experimental SAXS profiles with calculated triaxial ellipsoidal fits (grey) for the prepared ssDNA-AuNPs (top) and the fit residuals (bottom). Scattering curves are offset for clarity and colored as in panel (A). Numerical values are summarized in Table 4-4. (C) Polydispersity of the major equatorial radius (r_A) modeled as a Gaussian distribution using SasView. Plots are scaled to the calculated r_A values and colored as in panel (A). Numerical values for panels (A-B) are summarized in Table 4-4. (D) DLS histograms reveal increased hydrodynamic radii of AuNPs after ssDNA conjugation and mPEG-SH coating, as compared to diameters calculated from corresponding *P(r)* functions (Table 4-7). Note that all ssDNA-AuNPs are also PEGylated.
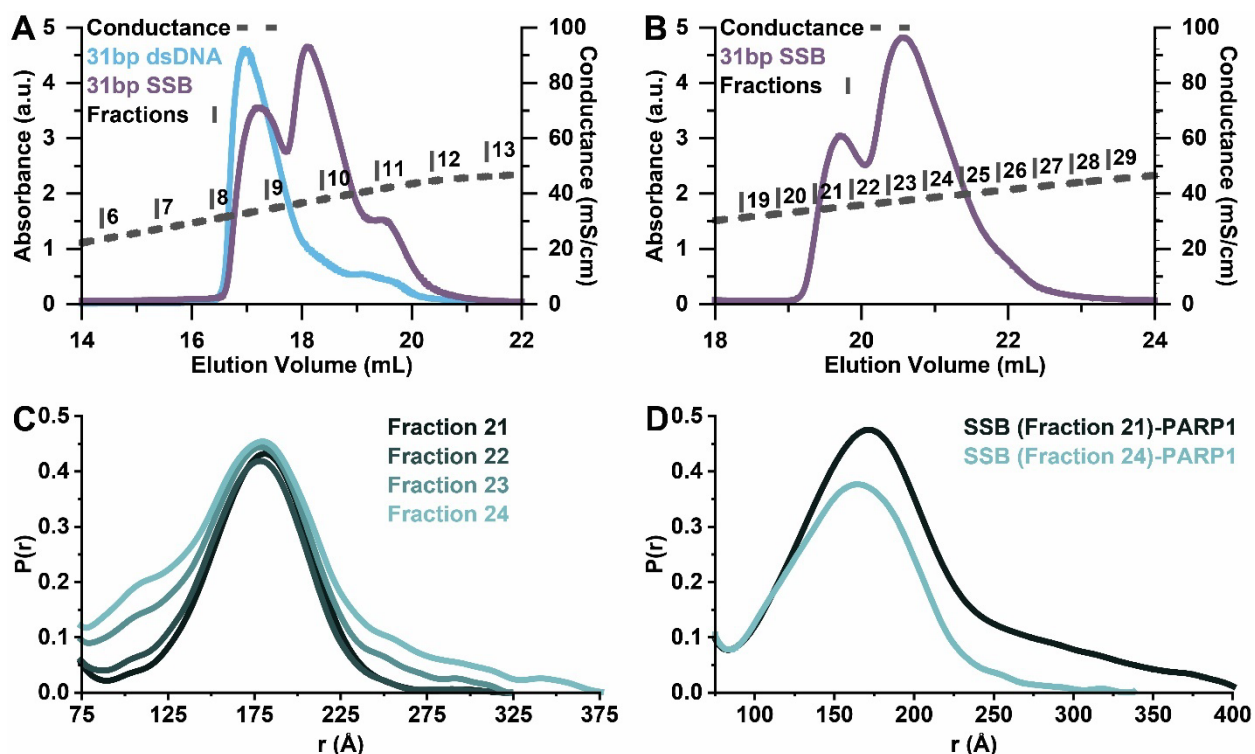
Figure 4-7. Anion exchange chromatograms for secondary purification of duplexed dsDNA without (blue) and with SSB (purple) for (A) rough and (B) refined fractionation. Spectra are measured by diode array detector (DAD) at 520 nm. Conductance measurements depict salt gradient conditions (dash grey). Fractionation shown by enumerated vertical lines. Pairwise distribution functions, *P(r)*, show inter-AuNP distance peaks for single-strand break (SSB) dsDNA substrate fractions enumerated in Panel B (C) alone and (D) in the presence of PARP-1. *P(r)* functions are normalized to the primary intra-AuNP peak, then this peak is omitted for clarity (not on the x-scale).

Table 4-4. Physical parameters of synthesized ssDNA-AuNPs obtained from pairwise distribution functions and triaxial ellipsoidal fits of scattering curves.

P(r) Peak Diameter (nm)	Major Equatorial Radius, r_A (nm)	Minor Equatorial Radius, r_B (nm)	Polar Radius, r_C (nm)	Polydispersity Index (PDI)	χ^2-value	r_A / r_C
5.75 ± 0.10	3.57 ± 0.09	2.90 ± 0.03	2.21 ± 0.02	0.25	1.25	1.6
6.92 ± 0.12	4.13 ± 0.07	4.13 ± 0.05	2.65 ± 0.01	0.20	0.31	1.6
7.38 ± 0.24	4.99 ± 0.16	3.50 ± 0.10	2.78 ± 0.03	0.10	0.19	1.8
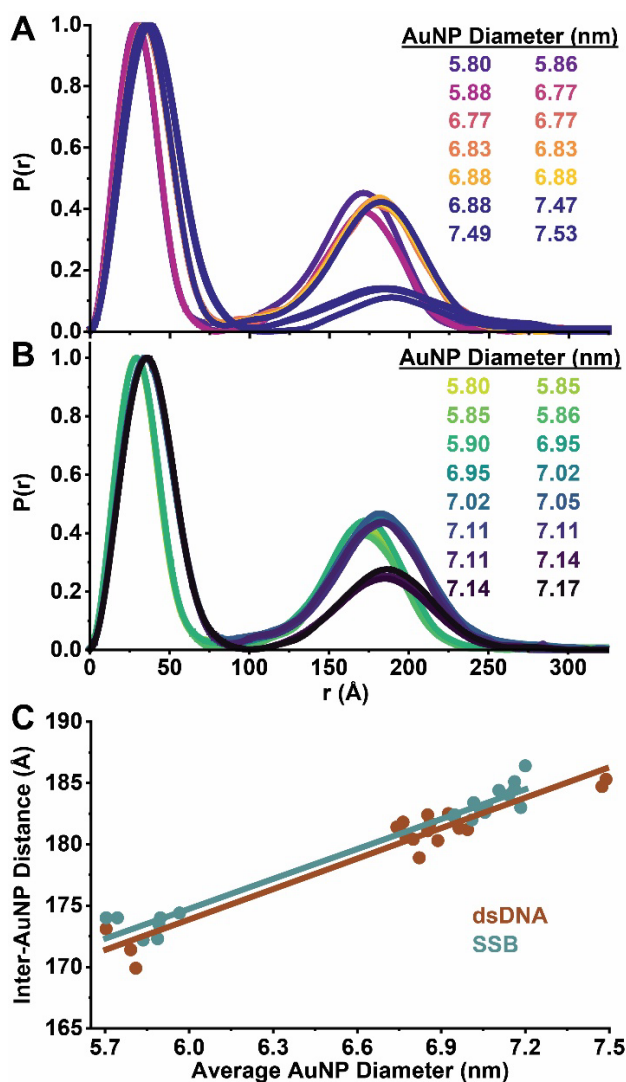
Figure 4-8. Demonstration of how the measured inter-AuNP distance and resolution changes as a function of AuNP size. (A-B) *P(r)* functions show inter-AuNP distance peaks for dsDNA substrates (A) without and (B) with a single-strand break (SSB) as a function of conjugated AuNP size. (C) Plot of inter-AuNP distance as a function of AuNP size over the diameter range of 5.7-7.5 nm.
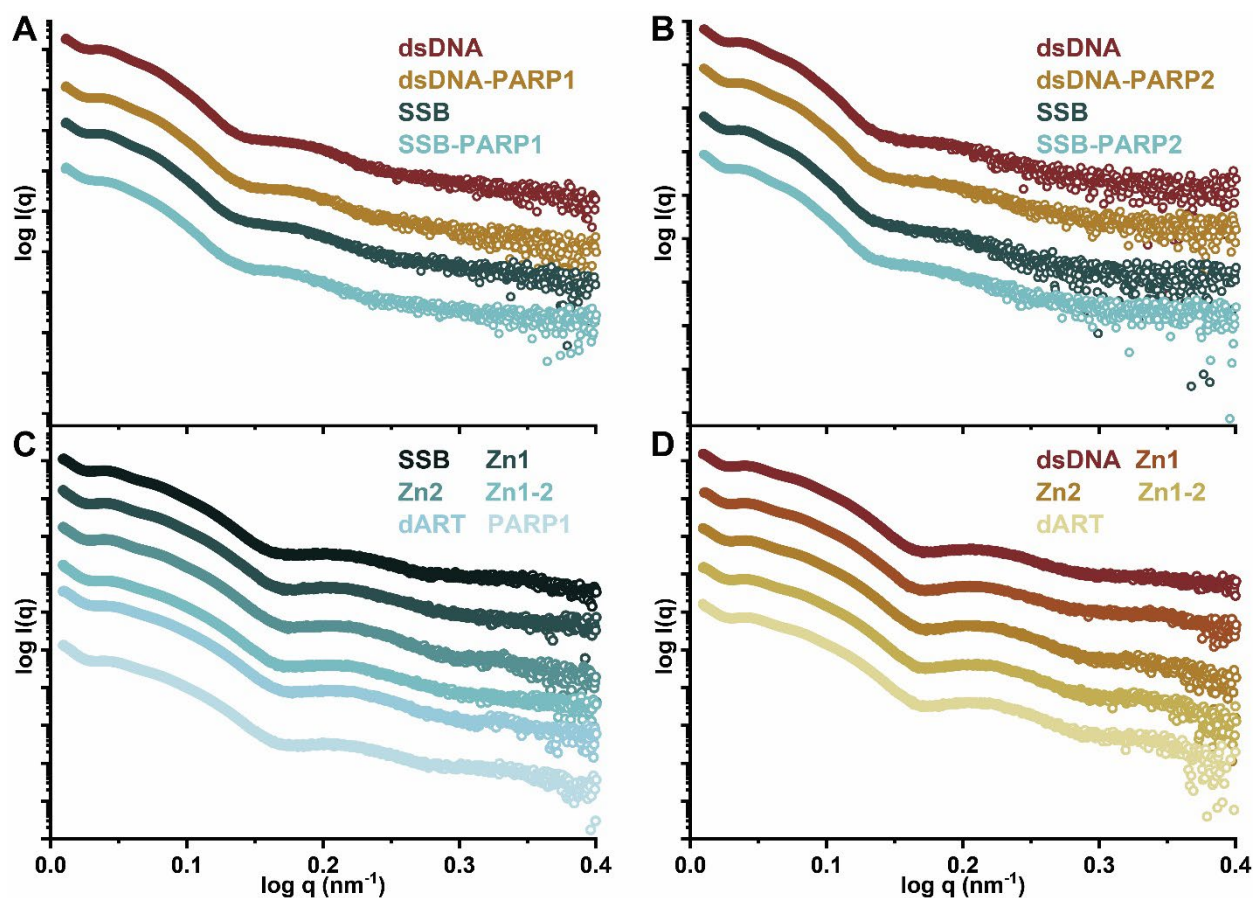
Figure 4-9. Scattering curves demonstrate an inter-AuNP distance shift only upon the interactions of PARP-1 with single-strand break (SSB) dsDNA and no movement is observed in the presence of PARP-2 or with fully duplexed dsDNA. (A-B) Scattering curves show inter-AuNP distance peaks for dsDNA substrates without (red series) and with a SSB (blue series) in the presence of either (A) PARP-1 and (B) PARP-2. (C-D) Scattering curves show inter-AuNP distance peaks for truncated versions of PARP-1 interacting with dsDNA substrates (C) with and (D) without a SSB. Scattering curves are offset for clarity.
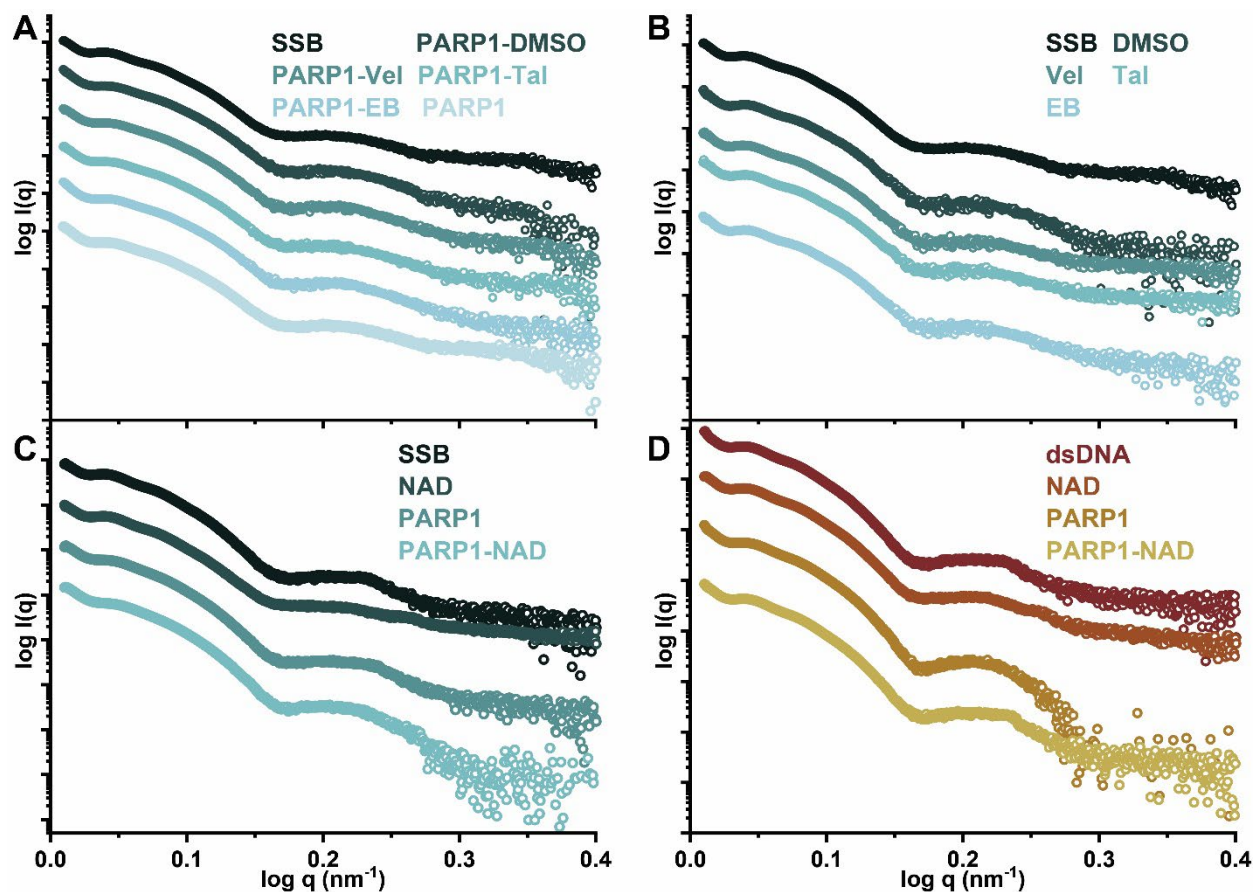
Figure 4-10. Scattering curves demonstrate that the inter-AuNP distance shift of PARP-1 interaction with SSB dsDNA remains unchanged by the presence of known PARP-1 inhibitors but shows a notably shift in the presence of NAD^+. Scattering curves show inter-AuNP distance peaks for (A-B) single-strand break (SSB) dsDNA substrates in the presence of known PARP-1 inhibitors Veliparib (Vel), Talazoparib (Tal), and EB-47 (EB), (A) with and (B) without PARP-1 and (C-D) PARP-1 interaction with (C) SSB dsDNA and (D) dsDNA in the presence of NAD^+. Scattering curves are offset for clarity.
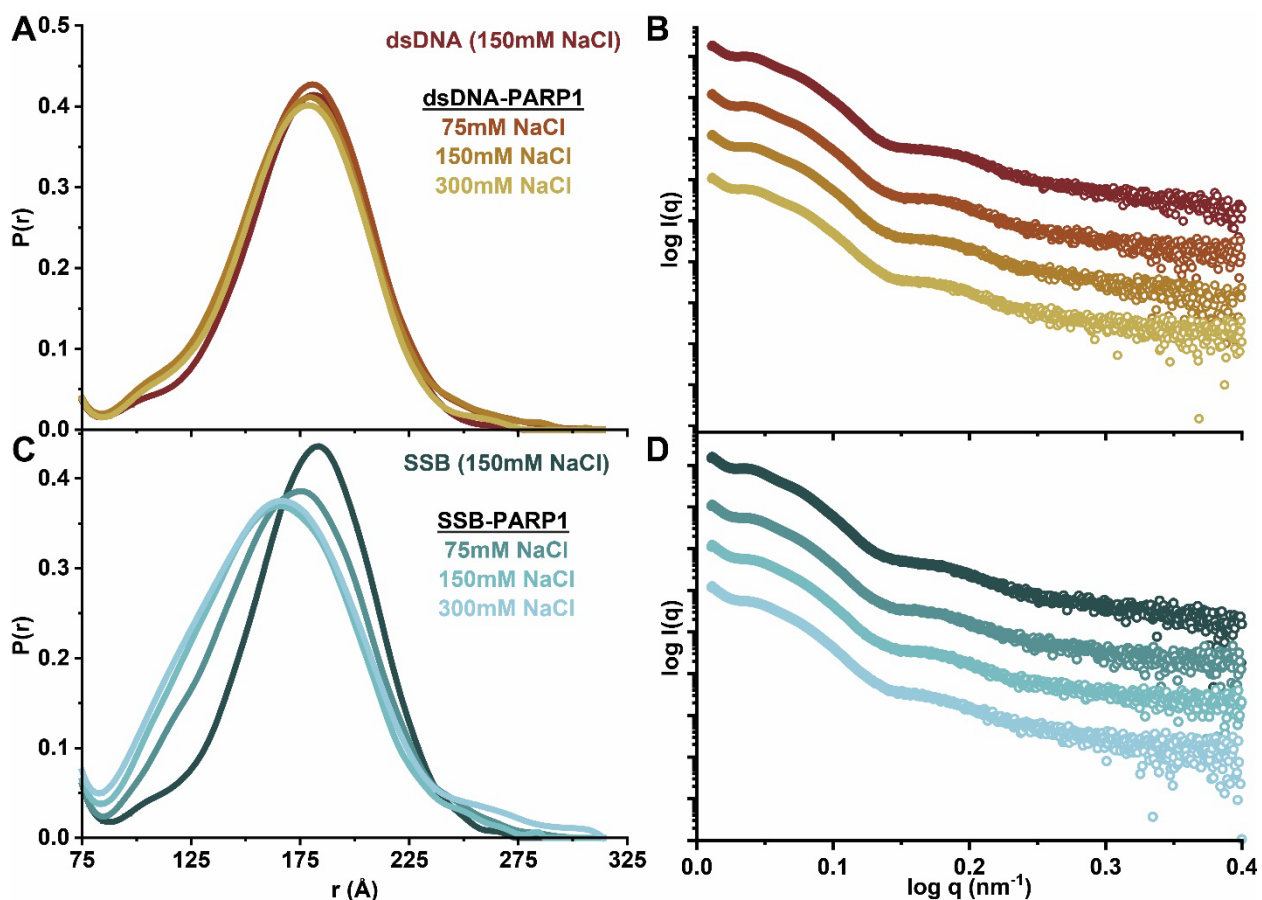
Figure 4-11. Pairwise distribution functions, *P(r)*, and scattering curves demonstrate that NaCl concentration does not affect PARP-1 with fully dsDNA but a minimum of 150mM NaCl is required for full binding of PARP-1 to dsDNA with a SSB. (A and C) *P(r)* functions and (B and D) show inter-AuNP distance peaks for PARP-1 interaction with dsDNA substrates (A-B) without and (C-D) with a single-strand break (SSB) as a function of NaCl concentration. *P(r)* functions are normalized to the primary intra-AuNP peak, then this peak is omitted (not on the x-scale) and scattering curves are offset for clarity.

67
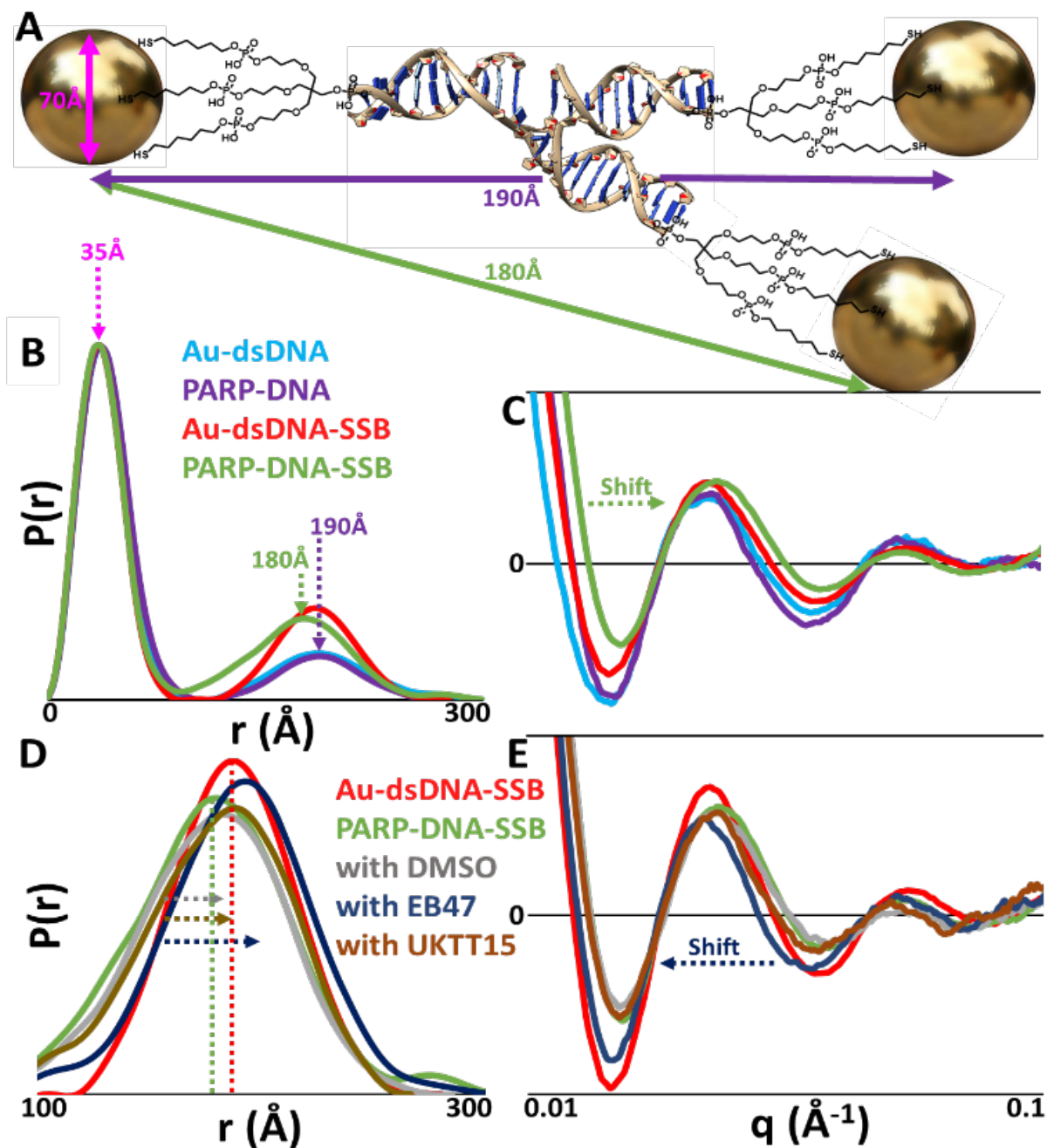
Figure 4-12. A) Schematic representation of substrate dynamics (not to scale). Preliminary data for dsDNA-AuNP substrates with and without a SSB, either alone or in the presence of PARP-1 shown as: B) distribution functions, $P(D_{i,j})$ and C) CSFs. Also, preliminary data showing CSFs for PARP-1 with added inhibitors in the presence of SSB dsDNA-AuNP substrates: D) distribution functions, $P(D_{i,j})$ and E) CSFs.

Table 4-5. Important Residues for PARP-1 binding and Activation.

Important Residues for PARP-1 DNA-binding and Activation*

Residue	Location	Function
Arg 34	Zn1	Backbone grip
Asp45	Zn1	Salt bridge to WGR
Arg78, Trp79	Zn1	Contact Zn3 Trp246 and Lys238 respectively
Trp246, Lys238	Zn3	Contact Zn1 Arg78 and Trp79 respectively
Trp318	Zn3	Fold central location at this interface WGR and HD
Asn567, Tyr569	WGR	Make additional contributions to the WGR–HD interface
Trp589	WGR	Stacks ribose sugars at the 5′-end of DSB, aids DNA exchange
Arg591	WGR	Provides a bridge between the Zn1 DNA damage interface and the CAT
Lys633	WGR	Contact Zn3 Trp318
Leu698, Leu701	HD	Leucine switch at hydrophobic core, α-helix C, distorted during binding DSB
Arg735	HD	Contact Zn3 Trp318
Glu988	ART	Form all interdomain contacts observed in the PARP-1/DNA structure, mutation partially restores HD destabilization of PARP-1

* Important residues information obtained from previous literature.[100,101]

Table 4-6. Important Residues for PARP-1 Inhibition.

Potential Residues Effected by PARP Inhibitor Binding*

Residue	Location	Interacting PARP Inhibitor	Function
Arg18	Zn1	olaparib	DNA interacting residue
Pro42	Zn1	talazoparib	DNA interacting residue
Phe44	Zn1	talazoparib	WGR interacting residue
Trp79	Zn1	niraparib	Zn3 interacting region
Ser274	Zn3	all	DNA interacting
Thr316	Zn3	all	CAT interacting residue
Val563	WGR	veliparib	Zn1 interacting residue
Ile562	WGR	veliparib	CAT interacting residue
Asn567, Ser568	WGR	olaparib	Zn1 interacting residue
Pro635	WGR	all	Zn3 interacting residue
Asp644	WGR	all	CAT interacting residue
Lys700, His742, Gly745, Lys747, Lys748	HD	all	WGR interacting residues
Glu763, Asp766	HD	niraparib, veliparib	forming water-mediated H-bonds
Asp766 and Asp770	HD	EB-47	EB-47 bound clashes
Gly863	ART	Olaparib, veliparib	side chains H-bond contacts
Ser904	ART	rucaparib, niraparib, talazoparib	side chains H-bond contacts
Tyr889	ART	talazoparib	specific π-π stacking interaction

* Potential residues information obtained from previous literature.[89,123,130,143,144]

Table 4-7. Physical parameters of synthesized AuNPs as obtained from pairwise distribution functions of X-ray scattering and dynamic light scattering.

	$P(r)$ Peak Diameter (nm)	DLS Peak Diameter (nm)	Increase in Diameter from PEGylation (nm)
PEG-AuNP	6.82 ± 0.03	10.33 ± 0.22	3.51 ± 0.25

5 Quantitative Protein Corona Composition, Driving Forces, and Dynamics on Carbon Nanotubes in Biological Environments[§]

5.1 Chapter Abstract

When nanoparticles enter biological environments, proteins adsorb to form the "protein corona" which alters nanoparticle biodistribution and toxicity. Herein, we measure protein corona formation on DNA-functionalized single-walled carbon nanotubes (ssDNA-SWCNTs), a nanoparticle used widely for sensing and delivery, in blood plasma and cerebrospinal fluid. We characterize corona composition by mass spectrometry, revealing high-abundance corona proteins involved in lipid binding, complement activation, and coagulation. We investigate roles of electrostatic and entropic interactions driving selective corona formation. Lastly, we study real-time protein binding on ssDNA-SWCNTs, obtaining agreement between enriched proteins binding strongly and depleted proteins binding marginally, while highlighting cooperative adsorption mechanisms. Knowledge of protein corona composition, formation mechanisms, and dynamics informs nanoparticle translation from *in vitro* design to *in vivo* application.

5.2 Introduction

Engineered nanoparticles are prominently used for sensing and imaging applications in biological systems due to their distinctive optical and physical properties.[157,158] A key challenge with these nanoscale tools is understanding the mechanisms of interaction between the nanoprobe and the biological system they are designed to query.[159,160] An incomplete understanding of protein corona formation remains as a paramount barrier to successfully implementing nanotechnologies within biological environments.

Although many studies classify protein corona composition around specific nanoparticle systems, significant debate persists as to which protein and nanoparticle characteristics are most important in determining corona composition, and how different biological environments contribute to compositional and temporal corona heterogeneity.[159,161] While prior studies clarify different aspects of bio-corona formation, system constraints such as surface-immobilization or treating the protein corona as existing at thermodynamic equilibrium make it difficult to reliably translate results to real biofluid systems.[160,162,163] Additionally, many nanosensor technologies are tested for biofouling and biocompatibility in blood serum, a blood-based fluid rich in albumin, the most abundant blood plasma protein, and devoid of blood clotting proteins. The assumptions that serum is a representative biofluid for confirming *in vivo* function and that protein abundance in a native biofluid determines its relative abundance in a nanoparticle corona both stand to be refined.

Understanding protein corona formation is essential to design nanoparticles that are robust and stable in biological environments. Our work focuses on single-walled carbon nanotubes

(SWCNTs), a nanoparticle class that possesses unique optical and physical properties ideal for biological imaging, molecular sensing, and delivery applications.[157,164-166] To apply hydrophobic SWCNTs in aqueous biological systems, noncovalent functionalization with amphiphilic polymers imparts water solubility to the SWCNT, while retaining the near-infrared-emissive electronic structure.[164] Select polymers confer molecular recognition functionality when adsorbed to the SWCNT surface, such as single-stranded DNA (ssDNA). Specifically, ssDNA sequences $(GT)_6$ or $(GT)_{15}$ adsorbed to SWCNTs are implemented to image the neurotransmitter dopamine in the brain at spatiotemporal scales of relevance to endogenous neuromodulation.[166-168] ssDNA-functionalized SWCNTs have further been applied in intravenous *in vivo* scenarios, to monitor endolysosomal lipid accumulation[169] and nitric oxide production.[170] To design and apply these and other SWCNT-based nanotechnologies in biological systems, it is crucial to understand the composition, dynamics, and dominant mechanisms of protein corona formation.

This original publication explores protein corona formation probed with a selective adsorption assay generalizable to different types of nanoparticles and biofluids. We focus on two nanoparticles: a model system of commonly studied polystyrene nanoparticles (PNPs)[162,171-174] and a newer system of noncovalently functionalized SWCNTs. Protein adsorption on these nanoparticles is assessed in two biofluids: blood plasma, a standard biofluid relevant for blood circulation applications, and cerebrospinal fluid (CSF), an understudied biofluid relevant for central nervous system studies. An understanding of the corona formed on SWCNTs in CSF has not been investigated and is imperative for developing SWCNT-based applications in the brain, including mapping of the brain extracellular space[175] and vasculature,[176] neurotransmitter imaging,[166,177] and delivery to the brain.[165] Corona composition characterized by quantitative, label-free mass spectrometry analysis reveals key protein corona contributors and isolation of protein factors governing corona formation. We identify interactions driving protein adsorption, where hydrophobic interactions dominate formation of the inner corona, while electrostatic interactions govern formation of the outer corona.

The content of this chapter has been abrigded to focus spacio-temporal corona formation. To quantify the time-dependent protein corona formation process, we assess binding thermodynamics and kinetics by measuring adsorption of key proteins to $(GT)_{15}$-SWCNTs via isothermal titration calorimetry (ITC) and a corona exchange assay.[178] The protein-SWCNT complex structure is ascertained by small-angle x-ray scattering (SAXS), demonstrating changing mass fractal morphology of ssDNA-SWCNTs in the presence of a high-binding protein (fibrinogen) otherwise absent with the low-binding protein (albumin). Overall, we present a holistic experimental approach and analysis methodology to understand the complexities of protein corona formation, and apply this framework to examine an understudied system of interest: SWCNT-based probes in the brain.

5.3   Results and Discussion

5.3.1   *Protein Corona Composition*

Protein corona composition was studied on $(GT)_{15}$-functionalized SWCNTs (see synthesis in section 5.5.1; average 1 nm diameter, 500 nm length) and PNPs (100 nm diameter) in blood plasma (normal human, pooled donors; Innovative Research Inc.) and cerebrospinal fluid (CSF;

Table 5-1. Top 20 most abundant proteins identified by proteomic mass spectrometry in plasma nanoparticle coronas.

	Plasma	PNPs in plasma	(GT)$_{15}$-SWCNTs in plasma
1	Serum albumin	Alpha-2-HS-glycoprotein	Clusterin
2	Haptoglobin	Ig kappa constant	Histidine-rich glycoprotein
3	Ig kappa constant	Haptoglobin	Apolipoprotein A-I
4	Ig heavy constant gamma	Complement C3	Complement C3
5	Serotransferrin	Kininogen-1	Haptoglobin
6	Apolipoprotein A-I	Ig heavy constant gamma 1	A disintegrin and metalloproteinase with thrombospondin motifs 12
7	Complement C4	Apolipoprotein A-II	Complement C1r subcomponent
8	Telomeric repeat-binding factor 2-interacting protein	tRNA-dihydrouridine(47) synthase [NAD(P)(+)]-like	Vitronectin
9	Alpha-1-antitrypsin	Beta-2-glycoprotein 1	Kininogen-1
10	Alpha-2-HS-glycoprotein	Vitronectin	Prothrombin
11	Apolipoprotein A-II	Serum albumin	C4b-binding protein alpha chain
12	Ig heavy constant alpha 1	Vitamin D-binding protein	Complement factor H
13	Integrin alpha-7	A disintegrin and metalloproteinase with thrombospondin motifs 12	Fibrinogen alpha chain
14	Alpha-2-macroglobulin	Hemopexin	Protein AMBP
15	Complement C3	Apolipoprotein A-I	Beta-2-glycoprotein 1
16	Complement C5	Ig lambda-like polypeptide 5	Apolipoprotein E
17	Hemopexin	Histidine-rich glycoprotein	Complement C1q subcomponent subunit B
18	Alpha-1-acid glycoprotein 1	Clusterin	Ig heavy constant gamma 1
19	Ig heavy constant mu	Alpha-1-antitrypsin	Ig J chain
20	Beta-2-glycoprotein 1	Serum paraoxonase/arylesterase 1	Galectin-3-binding protein

normal human, pooled donors; Lee Biosolutions). Selective adsorption of proteins onto nanoparticles was evaluated by (i) incubating nanoparticles with biofluid for 1 h, (ii) isolating protein-nanoparticle complexes by centrifugation, (iii) removing unbound proteins by washing, (iv) eluting bound proteins from nanoparticles with surfactant and reducing agent, and (v) characterizing proteins by two-dimensional polyacrylamide gel electrophoretic separation (2D PAGE) or liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figure 5-3; see methods in section 5.5.5).[171] Following workflow validation (Figure 5-3 and Figure 5-4), protein coronas were studied on these two distinct nanoparticle surfaces (PNPs and (GT)$_{15}$-SWCNTs) in two biofluids (plasma and CSF). PAGE analysis confirmed that proteins showed selective enrichment or depletion fingerprints on nanoparticles (Figure 5-5). More in-depth protein corona composition studies were subsequently undertaken by performing in-solution trypsin digestion of proteins eluted from nanoparticles, followed by protein characterization with label-free, quantitative LC-MS/MS. Analysis by LC-MS/MS provides (i) molar corona protein abundances via comparison to an internal standard and (ii) enrichment or depletion in each nanoparticle corona, relative to protein concentrations in the native biofluid (see details in section 5.5.6). The

Table 5-2. Top 20 most abundant proteins identified by proteomic mass spectrometry in CSF nanoparticle coronas.

	CSF	PNPs in CSF	(GT)$_{15}$-SWCNTs in CSF
1	Serum albumin	Cystatin-C	Complement C3
2	Transthyretin	Complement C3	Clusterin
3	Alpha-1-antitrypsin	Clusterin	Histidine-rich glycoprotein
4	Prostaglandin-H2 D-isomerase	Prostaglandin-H2 D-isomerase	Galectin-3-binding protein
5	Serotransferrin	Alpha-2-HS-glycoprotein	Apolipoprotein E
6	Cystatin-C	Collagen alpha-2(XI) chain	Prostaglandin-H2 D-isomerase
7	Alpha-1-acid glycoprotein 1	Beta-2-glycoprotein 1	Kininogen-1
8	Hemoglobin subunit alpha	Gelsolin	Apolipoprotein A-I
9	Ig heavy constant gamma 1	Serotransferrin	Vitronectin
10	Vitamin D-binding protein	Vitronectin	Transthyretin
11	Ceruloplasmin	Ig heavy constant gamma 1	Gelsolin
12	Hemopexin	Apolipoprotein E	Ig heavy constant gamma 1
13	Apolipoprotein E	Fibulin-1	Serotransferrin
14	Ig kappa constant	Major prion protein	Complement C1s subcomponent
15	Apolipoprotein A-I	Kininogen-1	Complement C1q subcomponent subunit B
16	Hemoglobin subunit beta	EGF-containing fibulin-like extracellular matrix protein 1	Fibulin-1
17	Haptoglobin	Complement factor H	Complement factor H
18	Clusterin	Histidine-rich glycoprotein	Major prion protein
19	Suppression of tumorigenicity 18 protein	Fibrinogen beta chain	Fibrinogen alpha chain
20	Gelsolin	ProSAAS	Cystatin-C

twenty most abundant proteins in the nanoparticle coronas are summarized in Table 5-1 (plasma) and Table 5-2 (CSF).

5.3.2 *Protein Corona Dynamics*

Beyond probing corona composition at the end point of adsorption, we investigated corona formation dynamics to understand the time-dependent process and overall system energetics driving corona formation. Toward this end, isothermal titration calorimetry (ITC) was applied to probe the thermodynamics of protein adsorption to SWCNT surfaces.[179-181] We studied binding of (GT)$_{15}$-SWCNTs with two proteins identified by LC-MS/MS with opposite binding affinities: albumin, selected as a model low-binding protein, and fibrinogen, a model high-binding protein. ITC results confirm that fibrinogen preferentially adsorbs to (GT)$_{15}$-SWCNTs and albumin does not, as evidenced by the binding curve in the former and absence of changing heats upon injection in the latter (Figure 5-11). From the ITC binding curve of fibrinogen with (GT)$_{15}$-SWCNTs, the change in enthalpy is -565.2 kJ/mol and the change in entropy is -1.756 kJ/K-mol.

This favorable enthalpic term outweighs the net unfavorable entropic terms to ultimately drive formation as a spontaneous, energetically favorable process: the net change in free energy is -41.91 kJ/mol. However, these ITC results must be interpreted with the consideration that the equilibrium requirement for this thermodynamic analysis is not rigorously held (see extended discussion in section 5.6.2).[160,162,163] This binding profile shape for protein-surface adsorption processes often emerges as a result of adsorption-induced protein spreading/denaturation, reorientation, and aggregation as functions of bulk protein concentration, in contrast to originating from the dynamic equilibrium between the fluid and surface-adsorbed phases required for Langmuirian adsorption.[182-184] Thus, although these binding curves confirm compositional findings of the relative binding affinities, it should be noted that ITC is not a suitable methodology to study all nanoparticle-protein systems and these limitations must be reflected in interpreting these energetics as overall changes in system energies, rather than a true deconvolution of protein-nanoparticle binding interactions.
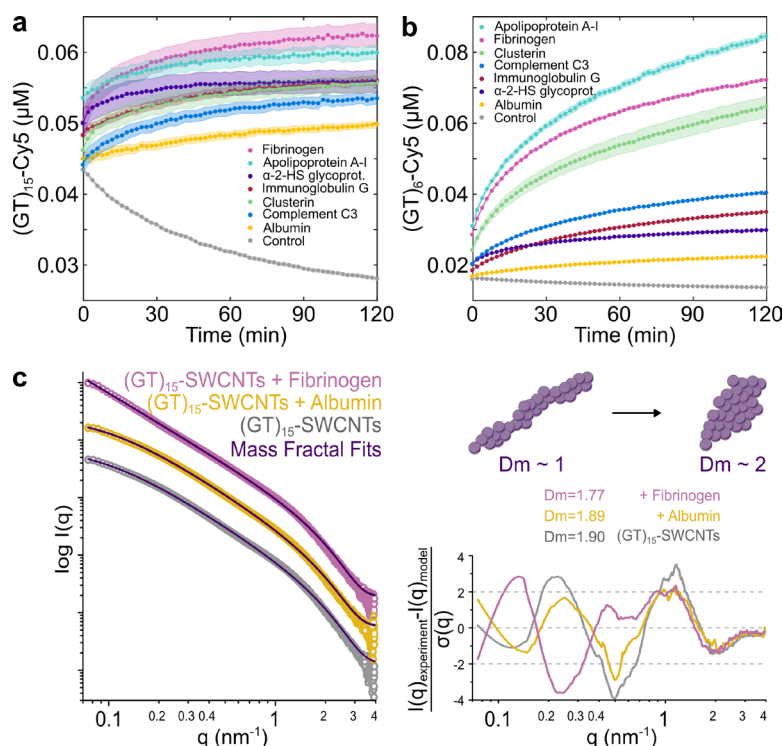
Figure 5-1. Protein corona dynamics and structure assessed for binding of key proteins to ssDNA-SWCNTs. A corona exchange assay is employed to determine binding kinetics of a protein panel (each at 80 mg L^{-1} final concentration) to (a) $(GT)_{15}$-SWCNTs and (b) $(GT)_6$-SWCNTs (each at 5 mg L^{-1} final concentration). Shaded error bars indicate standard error between experimental replicates (N = 3). Small-angle x-ray scattering (SAXS) is applied to gain in-solution structural information of albumin vs. fibrinogen adsorption on $(GT)_{15}$-SWCNTs. (c) Experimental SAXS profiles for 0.5 g L^{-1} $(GT)_{15}$-SWCNTs with and without albumin or fibrinogen, each at 0.5 g L^{-1} final concentrations. Mass fractal model fits are included in purple together with fit residuals on the right. The accompanying illustration depicts the mass fractal dimension Dm increasing from approximately 1 (rod-like) to 2 (disk-like), with the fit Dm values for $(GT)_{15}$-SWCNTs in the presence and absence of proteins.

We next implemented a real-time kinetic binding assay to study dynamic protein interactions with SWCNTs.[178] Briefly, multiplexed fluorescence enables tracking each entity involved in the corona formation process, with cyanine 5 (Cy5)-tagged ssDNA originally on the SWCNT surface exchanging with protein added to solution. We implemented this platform to track the binding of key plasma corona proteins to $(GT)_{15}$-SWCNTs and $(GT)_6$-SWCNTs (Figure 5-1a-b), with desorption of Cy5-tagged ssDNA originally on the SWCNT measured as an increase in Cy5 fluorescence and used as a proxy for protein adsorption to SWCNT. Specifically, we assayed the protein panel: clusterin, apolipoprotein A-I, fibrinogen, and complement C3, which are predicted to adsorb in high abundance to $(GT)_{15}$-SWCNTs, and alpha-2-HS glycoprotein, immunoglobulin G, and albumin, which are predicted to adsorb less to $(GT)_{15}$-SWCNTs based on LC-MS/MS compositional analysis (see expected ordering in Table 5-1). Interestingly, the order of protein adsorption from this corona exchange assay was: fibrinogen > apolipoprotein A-I > alpha-2-HS glycoprotein > immunoglobulin G ≈ clusterin > complement C3 > albumin (Figure 5-1a). While this result affirms the high affinity of fibrinogen and apolipoprotein A-I vs. low affinity of albumin to $(GT)_{15}$-SWCNTs, some of the single-protein end states do not match the relative ordering of protein abundances from the full-biofluid LC-MS/MS experiments. Accordingly, higher order interactions such as the Vroman effect are further supported in affecting protein adsorption in the full-biofluid experiments, absent in the single-protein experiments. Moreover, these time-dependent dynamics reveal that the rates of protein binding are distinct among proteins, even though some converge to the same final value (such as alpha-2-HS glycoprotein and clusterin). A comparison of this same protein panel binding to $(GT)_6$-SWCNTs is provided because the shorter ssDNA strand is displaced more readily, offering a greater spread between protein species (Figure 5-1b; see expected ordering in Table 5-5). The dynamics of protein adsorption recapitulate similar high- vs. low-binding propensities, yet, complement C3 and clusterin again display significantly less adsorption than expected based on LC-MS/MS results, signifying that these proteins enter the corona with cooperative binding mechanisms (e.g. C3 binding to other surface-adsorbed proteins[185]) rather than by high binding affinity to the SWCNT surface on their own. To build a physical picture of protein-SWCNT association, we next expand to structural studies of these protein-nanoparticle complexes.

5.3.3   Protein Corona Morphology

To evaluate in-solution structural changes of the $(GT)_{15}$-SWCNTs due to protein corona formation, small-angle X-ray scattering (SAXS) was performed with two proteins, albumin and fibrinogen, as low-binding and high-binding proteins, respectively. SAXS results confirm formation of unique form factors and thus complexation for $(GT)_{15}$-SWCNTs with fibrinogen, absent for the case of albumin (Figure 5-1c), therefore recapitulating corona compositional findings.

The intrinsically disordered experimental SAXS profiles were fit using mass fractal geometries, complemented by power-law dependencies from the Porod region, as detailed in extended discussion in section 5.6.3 (Figure 5-1c, Figure 5-12, and Table 5-8).[186-188] The mass fractal radii (all R~1 nm), traditionally defined as the radius of the uniform sphere used to cover the fractal, suggest that the overall topology of the $(GT)_{15}$-SWCNTs remains constant with and without protein. The fractal dimension Dm and analogous power-law exponent p, found to be in close

agreement, estimate the bulk geometries of the mass fractals, where the integer value represents the three dimensions in Euclidean space such that values of 1, 2, and 3 represent rod, disk, and sphere geometries, respectively. The decrease in fractal dimension from Dm~1.90 for $(GT)_{15}$-SWCNTs with or without albumin to Dm=1.77 for $(GT)_{15}$-SWCNTs with fibrinogen reveals an initial disk-like mass fractal geometry, then elongation to gain rod-like character in the presence of fibrinogen (Figure 5-1c). This is consistent with previous literature in which fibrinogen binds to SWCNTs in a lengthwise manner.[189,190] Furthermore, the decrease of $(GT)_{15}$-SWCNT Dm in the presence of fibrinogen signifies increasing attractive forces between the molecular entities and consequent colloidal instability.[191] Finally, the cutoff length ζ, or the maximum distance between any two points of the mass fractal, undergoes a ten-fold increase for $(GT)_{15}$-SWCNTs with fibrinogen, denoting a drastic increase in the aggregate size. Thus, SAXS confirms fibrinogen complexation with $(GT)_{15}$-SWCNTs, suggests a side-on orientation (as reiterated by TEM, Figure 5-13), and enables quantification of the changing fractal structure, pointing to the role of multilayer adsorption mechanisms and aggregate formation.

5.4   Conclusions

As engineered nanoparticles are increasingly implemented as tools to study and alter biosystems, it is crucial to develop an understanding of how these nanoparticles interact with their biological surroundings. Accordingly, we have conducted a multimodal study to characterize protein corona formation in a biologically representative in-solution state. We focus on applying $(GT)_{15}$-SWCNTs in the brain microenvironment, although the framework itself is generic to study protein corona composition on other nanoparticles and in other biofluids.

We find that while PNPs are largely agnostic to protein adsorption, $(GT)_{15}$-SWCNTs show strong preferential binding of proteins involved in lipid transport, complement activation, and blood coagulation. Importantly, enrichment of complement proteins (especially C3) on ssDNA-SWCNTs is concerning due to the potential of nanoparticle opsonization and complement pathway activation. The selectivity of proteins binding to $(GT)_{15}$-SWCNTs motivates either the rational design of sensors harnessing innate affinity for the SWCNT surface, or the development of SWCNT-based nanosensors passivated against detrimental biofouling. Additionally, $(GT)_{15}$-SWCNTs show high binding of fibrinogen and low binding of albumin, despite the prevalence of albumin binding on other nanoparticles across a body of previous literature.[192] This raises cogent concern for the need to test nanotechnologies in blood plasma (with all protein constituents present) rather than blood serum (absent of fibrinogen), where fibrinogen may be a more important contributor to diminished *in vivo* efficacy than albumin.

We connect protein attributes that dictate protein-nanoparticle interactions to the thermodynamics and transient kinetics of protein-nanoparticle binding. Outer corona formation can be mitigated by tuning electrostatic interactions through nanoparticle design and by applying dynamic flow conditions (such as in circulating environments), whereas entropic considerations must be considered for the inner corona. Moreover, protein properties mediate adsorption differently in each biofluid, underscoring the complexity of corona formation. This phenomenon emphasizes that protein corona formation is a function of collective interactions at the nano-bio interface, rather than a property of isolated protein-nanomaterial interactions.

This work clarifies fundamental interactions for nanoscale systems in which development and optimization is done *in vitro*, with a desired application *in vivo*. Difficulties persist in the effective application of ssDNA-SWCNTs in brain imaging and delivery, including biofouling and the tendency of ssDNA-SWCNTs to aggregate in the presence of proteins. A more in-depth understanding of the protein corona could allow *a priori* prediction of biodistribution profiles and/or enable us to better understand these results in organisms. Elucidating protein corona composition, dynamics, structure, and driving forces that mediate nanoparticle-protein interactions will establish design considerations for nanosensor development and provide a framework for understanding *how* and *why* our engineered nanoparticles are affecting, and being affected by, complex bioenvironments.

5.5   Materials and Methods

5.5.1   Synthesis of SWCNT-Based Nanosensors

Single-stranded DNA with single-walled carbon nanotube (ssDNA-SWCNT) suspensions were prepared with 1 mg of mixed-chirality SWCNTs (small diameter HiPco™ SWCNTs, NanoIntegris) and 1 mg of ssDNA (custom ssDNA oligos with standard desalting, Integrated DNA Technologies, Inc.) in 1 mL of 1X phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). Solutions were bath sonicated for 10 min (Branson Ultrasonic 1800) and probe-tip sonicated for 10 min in an ice bath (3 mm probe tip at 50% amplitude, 5-6 W, Cole-Parmer Ultrasonic Processor). Samples were incubated at room temperature for 30 min then centrifuged to pellet insoluble bundles and contaminants (16.1 krcf, 30 min). Supernatant containing the product was collected. ssDNA-SWCNTs were spin-filtered to remove free ssDNA (Amicon Ultra-0.5 mL centrifugal filters with 100 kDa MWCO, Millipore Sigma) by washing with Milli-Q water two times (8 krcf, 5 min) then reversing the spin filter and centrifuging to recover sample (1 krcf, 5 min). ssDNA-SWCNT concentration was determined via sample absorbance at 632 nm (NanoVue Plus, GE Healthcare Life Sciences) and the extinction coefficient $\varepsilon_{632nm}=0.036$ L mg^{-1} cm^{-1}.[164] ssDNA-SWCNTs were stored at 4°C until use and then diluted to a working concentration of 100 mg L^{-1} in PBS.

5.5.2   Nanoparticle Characterization

100 nm polystyrene nanoparticles were purchased from Polysciences, Inc. (PNPs; Fluoresbrite® yellow-green fluorophore-labeled). Size was confirmed with DLS to be in accordance with manufacturer specifications (Figure 5-3b) and measured zeta potential is -59.7 ± -3.24 mV (Figure 5-4; Zetasizer Nano, Malvern Panalytical; 1.67 g L^{-1} in PBS, 700 μL volume).

ssDNA-SWCNTs were synthesized as described above. Mixed-chirality HiPco™ SWCNTs are reported by the manufacturer (NanoIntegris) to have diameters 0.8-1.2 nm (average 1 nm; measured by Unidym from TEM) and lengths 100-1,000 nm (measured by Unidym from AFM). Upon suspension with ssDNA, previous AFM work informs a diameter of ~1 nm and length distribution centered around ~500 nm,[193] yet AFM sample deposition is known to influence such measurements in a DNA sequence-dependent manner.[194] Our previous work depicts ssDNA-SWCNT morphology by TEM.[195] Measured zeta potential of (GT)$_{15}$-SWCNTs is -19.4 ± 0.945

mV (Figure 5-4; Zetasizer Nano, Malvern Panalytical; 28.67 mg L^{-1} in PBS, 700 μL volume). Absorbance and fluorescence spectra of (GT)$_{15}$-SWCNTs are presented in Figure 5-2, confirming formation of a stable SWCNT dispersion[196] and in agreement with previous literature.[168] Absorbance of 30 mg L^{-1} (GT)$_{15}$-SWCNTs in PBS was measured in a 700 μL volume, black-sided quartz cuvettes (Thorlabs, Inc.) with a UV-VIS-nIR spectrophotometer (Shimadzu UV-3600 Plus). Fluorescence was obtained with an inverted Zeiss microscope (Axio Observer.D1, 10x objective) coupled to a Princeton Instruments spectrometer (SCT 320) and liquid nitrogen cooled Princeton Instruments InGaAs detector (PyLoN-IR). Fluorescence was measured in a glass-bottom 384 well-plate format (30 μL volume sample, 10 mg L^{-1} concentration in PBS), with a 721 nm laser (OptoEngine LLC) excitation light source and 800 – 1400 nm emission wavelength range.

We have previously determined approximately 140 (GT)$_{15}$ molecules or 364 (GT)$_{6}$ molecules adsorbed per SWCNT.[178] Using (GT)$_{15}$ and (GT)$_{6}$ contact areas from MD simulations,[168] this translates to ssDNA surface coverages of 2.1% and 6.5%, respectively. Previous work reports ~20-25% surface coverage of ssDNA on SWCNTs in the saturation regime (i.e. when further ssDNA adsorption is sterically unfavorable).[197-199] To capture the differing experimental and modeling conditions, we report the full range of ~1-25% initial ssDNA surface coverage on the SWCNT.

5.5.3   Isolation and Characterization of Protein-Nanoparticle Complexes

Protein corona composition was studied on PNPs, (GT)$_{15}$-SWCNTs, and (GT)$_{6}$-SWCNTs. PNPs were vortexed prior to use (1 min in 5 s pulses). Biofluids studied were human blood plasma and human CSF (Table 5-3) obtained with informed consent from all donors and in accordance with the relevant guidelines and regulations. CSF was concentrated 10X prior to incubation to match protein to nanoparticle ratios under volume constraints (14 krcf, 30 min; Amicon Ultra-0.5 mL centrifugal filters with 3 kDa MWCO, Millipore Sigma). The ratio of protein concentration to nanoparticle surface area was maintained constant for each respective nanoparticle in different biofluids, with 26 g L^{-1} protein per m^2 nanoparticle surface area for PNPs (from previous literature[171]) and 200 g L^{-1} protein per m^2 nanoparticle surface area for (GT)$_{6}$- and (GT)$_{15}$-SWCNTs. Based on experimental optimization, an 8-fold higher ssDNA-SWCNT surface area relative to PNP was selected to collect enough protein material from the SWCNT corona for downstream characterization, due to significantly lower protein adsorption on SWCNTs compared to PNPs. These incubation ratios translate to 1.67 g L^{-1} PNPs with 2.67% (v/v) plasma; 0.4 g L^{-1} PNPs with 8.67% (v/v) 10X CSF; 28.67 mg L^{-1} (GT)$_{15}$-SWCNTs with 2.67% (v/v) plasma; and 12.67 mg L^{-1} (GT)$_{15}$-SWCNTs with 16% (v/v) 10X CSF. Biofluid percentages are nominal and were adjusted on a mass basis to match the target protein per surface area ratios. Nanoparticles were incubated with biofluids in PBS, 750 μL total volume, for 1 h at ambient temperature (Figure 5-3a). Protein-nanoparticle complexes were pelleted by centrifugation (16.1 krcf, 20 min). Supernatant containing unbound proteins was removed, the pellet resuspended in PBS, and the pellet broken up by pipetting. Washing was repeated three times to ensure removal of unbound proteins.

Each step was validated for polystyrene nanoparticles (PNPs) exposed to blood plasma proteins as follows: (i) incubation of proteins with nanoparticles induced an increase in nanoparticle

hydrodynamic radius as determined by dynamic light scattering (DLS), where the number distribution shifted to a larger peak center and broadened out due to nonuniform aggregate formation as protein to nanoparticle loading was increased (Figure 5-3b); (ii) proteins initiated nanoparticle aggregation, as shown by solution absorbance before and after initial pelleting (Figure 5-3c), thus facilitating nanoparticle recovery for analysis; (iii) three washing steps were sufficient to remove unbound proteins by quantifying proteins remaining in the supernatant (Figure 5-3d; also valid for all nanoparticle/biofluid combinations); and (iv) proteins were fully eluted from nanoparticles by boiling in solutions of sodium dodecyl sulfate/β-mercaptoethanol (SDS/βME, for 2D PAGE analysis; Figure 5-3e) and urea/dithiothreitol (urea/DTT, for LC-MS/MS analysis). The equivalent verification was performed with (GT)$_{15}$-SWCNTs, yet the high aspect ratio of SWCNTs precluded accurate DLS measurement. Zeta potentials of the nanoparticle/plasma mixtures were determined as a proxy of the nanoparticle-protein complex surface charge, although this measurement captures a convolution of any free proteins, free nanoparticles, and nanoparticle-protein complexes (Figure 5-4). Zeta potential measurements of plasma proteins alone and nanoparticles alone reveal that the separate entities were initially negatively charged, whereby mixing results in a broadened zeta distribution of lower average magnitude than the nanoparticles alone. The measured reduction in effective surface charge implies some degree of protein adsorption to the nanoparticles and lowering of electrostatic repulsion, contributing to the experimentally observed colloidal instability upon combining nanoparticles with plasma, in agreement with previous literature.[172]

Toward (i), the incubation solution was characterized by dynamic light scattering and zeta potential measurements in folded capillary zeta cell disposable cuvettes (Zetasizer Nano, Malvern Panalytical; 700 µL volume). PNPs are negatively charged as a result of initiator fragments from the polymerization process, yet these PNPs are conventionally considered to be a model plain nanoparticle due to no explicit functionalization.[171] (GT)$_{15}$-SWCNTs are slightly negatively charged due to the presence of the ssDNA on the surface, with the phosphate backbone extending into solution. Toward (ii), absorbance spectra were measured in a 700 µL volume, black-sided quartz cuvettes (Thorlabs, Inc.) with a UV-VIS-nIR spectrophotometer (Shimadzu UV-3600 Plus). For (iii), free protein remaining in the supernatant after centrifugation was quantified during subsequent wash steps using the Qubit Protein Assay (Thermo Fisher Scientific). Note that PNPs contribute minimally to the Qubit signal (~2%), therefore the protein mass calculated for wash 0 is slightly inflated. For (iv), eluted protein from the nanoparticle was quantified using the Pierce 660nm Assay (with Ionic Detergent Compatibility Reagent; Thermo Fisher Scientific). Elution buffer was modified from SDS/βME for 2D PAGE to urea/DTT for LC-MS/MS analysis due to SDS interference with trypsin digestion, reverse-phase HPLC, and electrospray ionization efficiency.[200] The profile of eluted proteins was confirmed to be invariable to the elution system by 2D PAGE and S-trap (Protifi) LC-MS/MS analysis, although total eluted protein amount decreased.

Nanoparticle mass loss during pelleting and washing was estimated by measuring solution absorbance of each collected supernatant after centrifugation. This measured mass loss serves as a maximum estimate due to scattering of solubilized proteins and any remaining protein-nanoparticle aggregates that increase the absorbance baseline and impede fully accurate quantification of the nanoparticles alone. After each centrifugation step as shown in Figure 5-3a (four total), the supernatant was removed, and absorbance was measured (NanoDrop™

One/OneC Microvolume UV-Vis Spectrophotometer). For PNPs, absorbance was measured at the excitation maximum of the fluorophore (441 nm) and a standard curve over the relevant absorbance range (linear fit, R^2 = 0.9986) was used to convert this to concentration using Beer-Lambert's Law. For (GT)$_{15}$-SWCNTs, absorbance was measured at 632 nm and the known extinction coefficient was applied similarly. Results are presented in Table 5-4, with standard deviations of technical triplicate measurements and "0" denoting absorbance reading at the noise level of the instrument (e.g., absorbance ≤ 0.0133, read for buffer). The mass loss percentage is calculated as the ratio of this measured total mass removed to the calculated initial mass added to solution. We conclude that the maximum mass loss estimates of ~12% for PNPs and ~32% for (GT)$_{15}$-SWCNTs in each biofluid are not a significant portion of the population.

As a control, in the absence of nanoparticles in the incubation step, no measurable protein was present after pelleting and denaturation, confirming that we are measuring selective protein adsorption to nanoparticles, not merely to the container, nor simply seeing the high background of proteins in biofluids. This latter point is further confirmed by the result that protein corona abundance does not scale as a function of native abundance on ssDNA-SWCNTs (Figure 5-10). Contamination of the isolated protein corona with bio-nanoparticles, such as extracellular vesicles and lipoproteins,[201] was inferred by the aforementioned control (no "protein corona" measurable in the absence of nanoparticles) and the absence of large peaks in the plasma-alone DLS (Figure 5-3b).

5.5.4 Composition Studies by Two-Dimensional Polyacrylamide Gel Electrophoretic Separation (2D PAGE)

2D PAGE was performed to identify proteins via separation by isoelectric point in the first dimension and molecular weight in the second dimension. For analysis by 2D PAGE, bound proteins were eluted from nanoparticles by heating at 95°C for 10 min in SDS/BME reducing buffer (2% SDS, 5% β-mercaptoethanol, 0.066 M Tris-HCl). 1D separation was run according to the O'Farrell protocol[202] (adapted for Bio-Rad Mini-PROTEAN Tube Cell). Briefly, 1D sample buffer (8 M urea, 2% Triton X-100, 5% β-mercaptoethanol, 2% total carrier ampholytes - 1.6% Bio-Lyte 5/7, 0.4% Bio-Lyte 3/10) was added to samples in a 1:1 or 0.07:1 volume ratio (relative to initial plasma and CSF volumes, respectively) and incubated for 10 min. 1D separation was carried out in capillary tube PAGE with gel composition of 4% acrylamide (total monomer), 8 M urea, 2% Triton X-100, 2% total carrier ampholytes, 0.02% ammonium persulfate (APS), and 0.15% Tetramethylethylenediamine (TEMED). 25 μL sample and 25 μL 1D sample overlay buffer (4 M urea, 1% total carrier ampholytes) was loaded per capillary tube gel. The upper and lower chamber buffers were 100 mM sodium hydroxide and 10 mM phosphoric acid, respectively. 1D separation was run at 500 V for 10 min, 750 V for 3.5 h. Nanoparticles were filtered from the eluted proteins by the gel itself. Capillary gels were extruded and loaded onto 2D gels. 2D separation was run according to the Laemmli protocol[203] (adapted for Bio-Rad Mini-PROTEAN Tetra Cell). Briefly, SDS/BME reducing buffer was added to the 2D well to cover the capillary gel and incubated for 10 min. 2D separation was carried out in 1 mm vertical mini gel format with a discontinuous buffer system under denaturing conditions. Gel composition was 12% acrylamide (total monomer), 0.375 M Tris-HCl, 0.1% SDS, 0.05% APS, 0.05% TEMED for the resolving gel and 12% acrylamide (total monomer), 0.125 M Tris-HCl, 0.1% SDS, 0.05%

APS, 0.1% TEMED for the stacking gel. The electrode buffer was 25 mM Tris, 192 mM glycine, and 3.5 mM SDS (pH 8.3). 2D separation was run at 200 V for 1 h. Gels were extracted and silver stained according to Bio-Rad's Silver Stain Plus protocol and identified with ExPASy's SWISS-2DPAGE database (Figure 5-5).[204]

5.5.5 Composition Studies by Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

Bound proteins were eluted from nanoparticles by heating at 37°C for 60 min in urea/DTT reducing buffer (8 M urea, 5 mM DTT, 50 mM Tris-HCl, pH 8). Eluted protein concentration was determined with the EZQ Protein Quantitation Kit (Thermo Fisher Scientific). Protein solution was centrifuged to pellet the majority of nanoparticles (16 krcf, 20 min) and this supernatant was spin-filtered to concentrate and remove impurities (14 krcf, 30 min; Amicon Ultra-0.5 mL centrifugal filters with 3 kDa MWCO, Millipore Sigma; pre-rinsed). Proteins were alkylated with 15 mM iodoacetamide for 30 min in the dark. 500 mM DTT was added to quench excess iodoacetamide in a volume ratio of 3:1 and incubated for 20 min. The reaction was diluted 1:1 with 50 mM Tris-HCl pH 8 to allow enzymatic protein digestion. In-solution protein digestion was done with a ratio of 1:25 w/w Trypsin/Lys-C (Mass Spectrometry Grade, Promega) to protein, overnight at 37°C. Any remaining nanoparticles were removed by spin filtering (14 krcf, 30 min; Amicon Ultra-0.5 mL centrifugal filters with 30 kDa MWCO, Millipore Sigma; pre-rinsed). Nanoparticle removal was done after protein digestion into peptides due to the otherwise very similar sizes of nanoparticles and proteins. Peptide concentration was determined with the Pierce Peptide Quantitation Kit (Thermo Fisher Scientific) and samples were normalized to 0.1 g L^{-1} in 100 μL total volume. Peptide solutions were spiked with 50 fmol of E. coli housekeeping peptide (Hi3 Ecoli Standard, Waters) per 5 μL sample volume to allow for protein quantification. Digestion was terminated by freezing samples to -20°C. Note that biofluid-alone samples underwent these same processing steps, from denaturation to tryspin digestion. The preceding isolation steps of pelleting and washing were only necessary for nanoparticle-protein complexes and were accordingly omitted for biofluids alone. An alternative mass spectrometry preparation technique was pursued, using S-traps (Protifi), confirming our results were not biased by the sample preparation protocol.

Proteolytically digested proteins were analyzed using a Synapt G2-S*i* mass spectrometer equipped with a nanoelectrospray ionization source and connected directly in line with an Acquity M-class ultra-performance liquid chromatography system (UPLC; Waters, Milford, MA). This instrumentation is in the California Institute for Quantitative Biosciences (QB3)/College of Chemistry Mass Spectrometry Facility at UC Berkeley. Data-independent, ion mobility-enabled mass spectra and tandem mass spectra[205-207] were acquired in the positive ion mode. Data acquisition was controlled with MassLynx software (version 4.1) and tryptic peptide identification and quantification using a label-free approach[208-210] were performed with Progenesis QI for Proteomics software (version 4.0, Waters).

5.5.6   Proteomic Mass Spectrometry Data Interpretation

Prior to LC-MS/MS analysis, all samples were normalized on a total protein mass basis (where normalizing on a total molar basis is experimentally not feasible due to the complexity of biofluid samples). Consequently, the reported abundance of each protein species i, b_i, is the ratio of mole number of protein i, n_i, to the total protein mass:

$$b_i = \frac{n_i}{\sum_j n_j MW_j} \qquad\qquad 5\text{-}1$$

where MW_j is the molecular weight of each protein species j. LC-MS/MS data is then expressed as the fold change ε_i between the abundance of protein species i in the corona on the nanoparticle surface (phase s) to that in the bulk biofluid (phase f):

$$\varepsilon_i = \frac{b_i^s}{b_i^f} = \left(\frac{n_i^s}{n_i^f}\right)\left(\frac{\sum_j n_j^f MW_j}{\sum_j n_j^s MW_j}\right) \qquad\qquad 5\text{-}2$$

Here, the second term in parentheses is equal to 1 because all samples have the same total protein mass. Therefore, the reported fold change is the molar abundance ratio of a particular protein in the corona phase to that in the bulk biofluid phase.

5.5.7   Linear Regression Models for Corona Composition

We linearly regressed the natural log of the fold change of proteins for each nanoparticle-biofluid pairing using two sets of protein descriptors. The first set of descriptors are categorical variables denoting what class a protein is in (*i.e.* 1 for a protein in a given class and 0 otherwise), namely, involved in acute-phase response, blood coagulation, cell adhesion/signal transduction, complement activation, immune response, lipid binding/transport, regulation of biological processes, transport, or miscellaneous/unknown (Figure 5-6 and Figure 5-7; grouped according to PANTHER[211]). The variables were sum-effect coded such that the coefficients quantify how a protein class deviates from the grand mean of all protein classes and the intercept of the regression is the grand mean. Because each protein is grouped into one and only one class, the categorical variables are not linearly independent, and one class is excluded from the regression; we chose the miscellaneous class.

The second set of descriptors are molecular and biophysical properties of the proteins: protein mass, fraction of amino acids that are non-aromatic hydrophobic (sum of alanine, valine, isoleucine, leucine, and methionine content), hydrophilic (sum of serine, threonine, asparagine, glutamine content), arginine (R), histidine (H), lysine (K), acidic (sum of aspartic acid and glutamic acid content), phenylalanine (F), tyrosine (Y), tryptophan (W), number of glycosylated sites, number of ligand binding sites, number of metal binding sites, and number of disulfide binds (Figure 5-8 and Figure 5-9). Each of these descriptors is a continuous variable. The regression coefficients quantify the fractional difference in the fold change for a unit increase in the independent variable. Protein-specific information was acquired from UNIPROT.[212] Note that these particular descriptors were chosen after primary analyses that eliminated highly co-dependent descriptors. An example was choosing to include percentage of acidic/basic amino acids rather than protein isoelectric point (from ExPASy Compute pI/MW), where the isoelectric

point was deemed less exact because it relies on a theoretical calculation, omits protein fragments, and necessitates an average value for multicomponent proteins. Other examples were including number of disulfide bonds as an estimate of protein stability rather than protein instability index and segmenting to percentage of hydrophobic/aromatic amino acids rather than grand average hydropathy (GRAVY) score, in both cases due to the involvement of arbitrarily set scales (from ExPASy ProtParam).

For each regression, we included the measured protein fold changes for each replicate of a nanoparticle-biofluid system and controlled for sample-to-sample variability by including a categorical variable for the specific replicate. Protein abundances that fell below the lower limit of detection in the samples from the protein corona were set to $1x10^{-5}$ fmol, corresponding to the lowest detected protein abundance of all systems. Left-censoring the data in this way provides a conservative estimate of the regression coefficients by underestimating the magnitude and significance. Calculated variance inflation factors for all variables in each independent regression was <4, indicating negligible multicollinearity between the independent variables. To avoid overestimating the statistical significance of independent variables, p-values were adjusted using the Benjamini-Hochberg false discovery rate procedure. All statistical analysis was implemented in Python using the StatsModels V0.10.1 package (0.27-0.39). Table 5-6 and Table 5-7 provide coefficients, standard errors, false discovery rate corrected (FDRC) p-values, and R-squared values for each regression. The median R-squared of the first and second regression models for the nanoparticle-biofluid systems are 0.29 and 0.34, respectively, indicating the statistical models are descriptive rather than predictive. Nonlinear or decision tree algorithms provide more precise prediction of corona composition,[213] however, these approaches were not considered because they are not readily interpretable, which is a principle goal of our analysis.

Protein properties that were controlled for but that did not show a statistically significant effect on fold change for any nanoparticle in any biofluid include: the number of disulfide bonds (used as a proxy for protein stability), number of biomolecular binding sites, number of metal binding sites, and percentage of histidine or tryptophan. The lack of dependence on disulfide bond content and also instability index is surprising in the context of previous corona literature, which suggests that less structurally stable proteins are more surface active.[163]

5.5.8   Isothermal Titration Calorimetry (ITC) Methods

ITC measurements were performed with a NanoITC (TA Instruments). Prior to each experiment, samples and buffer were degassed for 10 min and the reference cell was filled with fresh Milli-Q water. Equilibration time was set to 1 h before the experiment started and the initial and final baselines were collected for 300 s. For each experiment, 1.2 g L^{-1} protein in PBS was titrated from the syringe (250 μL total volume) into 0.1 g L^{-1} (GT)$_{15}$-SWCNTs in PBS in the cell (1 mL total volume) under constant stirring (250 rpm) at 25 °C. 10 μL of protein titrant was injected into the nanoparticle solution in the cell every 7 min, with a total of 24 injections. By standard practice, every run was initiated with a 5 μL injection to ensure no artifacts due to bubbles and was removed from analysis. All protein-nanoparticle binding experiments were accompanied by three heat-of-dilution control experiments: (1) protein injected into buffer, (2) buffer injected into nanoparticles, and (3) buffer injected into buffer (where buffer is PBS). Heat of binding of protein to nanoparticles was then calculated as: (heat from titration of protein into nanoparticles)

– (1) – (2) + (3). Data processing was completed with NanoAnalyze software (TA Instruments). Baseline correction was done using the auto-fit routine. An independent binding model was applied to fit the fibrinogen data set, suitable to model weak nonspecific interactions such as those present in the system under study,[181] and a blank (constant) model was applied to fit the albumin data set.

Protein and nanoparticle concentrations and ITC setup parameters were varied in attempt of obtaining binding curves for both proteins to $(GT)_{15}$-SWCNTs. However, for albumin this was not possible within the ITC instrument's operational range, therefore albumin was concluded to not bind to $(GT)_{15}$-SWCNTs.

5.5.9   Corona Exchange Assay

Corona dynamic studies were completed as described previously.[178] Briefly, the same suspension protocol was employed for preparation of fluorophore-labeled ssDNA-SWCNT complexes, using ssDNA-Cy5 (3' Cy5-labeled custom ssDNA oligos with HPLC purification, Integrated DNA Technologies, Inc.) in place of unlabeled ssDNA. Lyophilized proteins were purchased (see details in Table 5-3) and reconstituted by adding 5 mg to 1 mL of PBS, tilting to dissolve for 15 min, filtering with 0.45 μm syringe filter (cellulose acetate membrane, VWR International), and quantifying with the Qubit Protein Assay (Thermo Fisher Scientific). Because of variation in amine-labeling of proteins, fluorescently labeled ssDNA was solely tracked, and the displacement of ssDNA from the SWCNT surface was taken as a proxy for protein adsorption. Equal volumes of 10 mg L^{-1} $(GT)_{15}$- or $(GT)_6$-Cy5-SWCNTs and 160 mg L^{-1} protein were added to a 96-well PCR plate (Bio-Rad) to a total volume of 50 μL. The plate was sealed with an optically transparent adhesive seal (Bio-Rad) and spun down on a benchtop centrifuge. Fluorescence time series readings were taken in a Bio-Rad CFX96 Real Time qPCR System, scanning the Cy5 channel every 2 min at 22.5°C. Fluorescence time series were analyzed without default background correction. Fluorescence values were converted to mass concentration using linear standard curves for ssDNA-Cy5. Note that in the case of the control, ssDNA adsorption to the SWCNT is observed, in line with previous studies.[178]

5.5.10 Structure Studies by Small-Angle X-ray Scattering (SAXS)

SAXS data was collected at SIBYLS beamline (bl12.3.1) at the Advanced Light Source of Lawrence Berkeley National Laboratory, Berkeley, California.[2] X-ray wavelength was set at $\lambda = 0.1127$ nm and the sample-to-detector distance was 2.1 m, resulting in scattering vector (q) ranging from 0.1–4 nm^{-1}. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, with scattering angle 2θ. Data was collected using a Dectris PILATUS3X 2M detector at 20°C and processed as described previously.[10]

Immediately prior to data collection, 30 μL of each sample was added to 96-well plates kept at 10°C and transferred to the sampling position via a Tecan Evo liquid handling robot with modified pipetting needles acting as sample cells as described previously.[43] Samples were exposed to X-ray synchrotron radiation for 30 s at a 0.5 s frame rate for a total of 60 images. Each collected image was circularly integrated and normalized for beam intensity to generate a

one-dimensional scattering profile by beamline-specific software. Buffer subtraction was performed for the one-dimensional scattering profile of each sample using each of two PBS buffer wells to ensure the subtraction process was not subject to instrument variations. Scattering profiles over the 30 s exposure were sequentially averaged together to eliminate any potential radiation damage effects. Averaging was performed with web-based software FrameSlice (sibyls.als.lbl.gov/ran).

5.5.11 Transmission Electron Microscopy (TEM) Methods

Holey carbon-coated grids (EMS Electron Microscopy Science) were surface-treated by glow discharge to make the support hydrophilic. Samples of $(GT)_{15}$-SWCNTs with fibrinogen or plasma were negatively stained with 1% uranyl acetate solution. For the $(GT)_{15}$-SWCNTs alone sample, no negative staining was done. 5 μL of 10 mg L^{-1} solution was drop-cast onto the grid. FEI ThemIS 60-300 STEM/TEM (National Center of Electron Microscopy, Molecular Foundry) with acceleration voltage of 60kV was used to acquire TEM images by video recording (Figure 5-13). A low acceleration voltage was chosen to minimize sample damage and increase sample contrast.

Table 5-3. Purchased biofluid and protein specifications.

Protein	Manufacturer	Lot #	Source	Form
Blood plasma	Innovative Research Inc.	#23791	Pooled normal human plasma	Biofluid
Cerebrospinal fluid	Lee Biosolutions	#07C5126	Pooled normal human CSF, from remnant lumbar puncture	Biofluid
Albumin	Sigma-Aldrich	#SLBZ2785	Human plasma	Lyophilized
Alpha-2-HS glycoprotein	Biovision Inc.	#4C08L75480	Human plasma	Lyophilized
Apolipoprotein A-I	Alfa Aesar	#927J17A	Human plasma	1 g L^{-1} in 10mM ammonium bicarbonate buffer, pH 7.4
Clusterin	R&D Systems	NEV1519031	Mouse myeloma cell line, NS0-derived human; Asp23-Arg227 (beta) & Ser228-Glu449 (alpha) with a C-terminal 6-His tag	Lyophilized
Complement C3	Mybiosource Inc.	#N30/20170	Human plasma	5 g L^{-1}
Fibrinogen	Millipore Sigma	#3169957	Human plasma	Lyophilized
Immunoglobulin G	Lee Biosolutions	#06B2334	Human plasma	Lyophilized

5.6 Chapter Supporting Information

5.6.1 Supplementary Figures and Tables
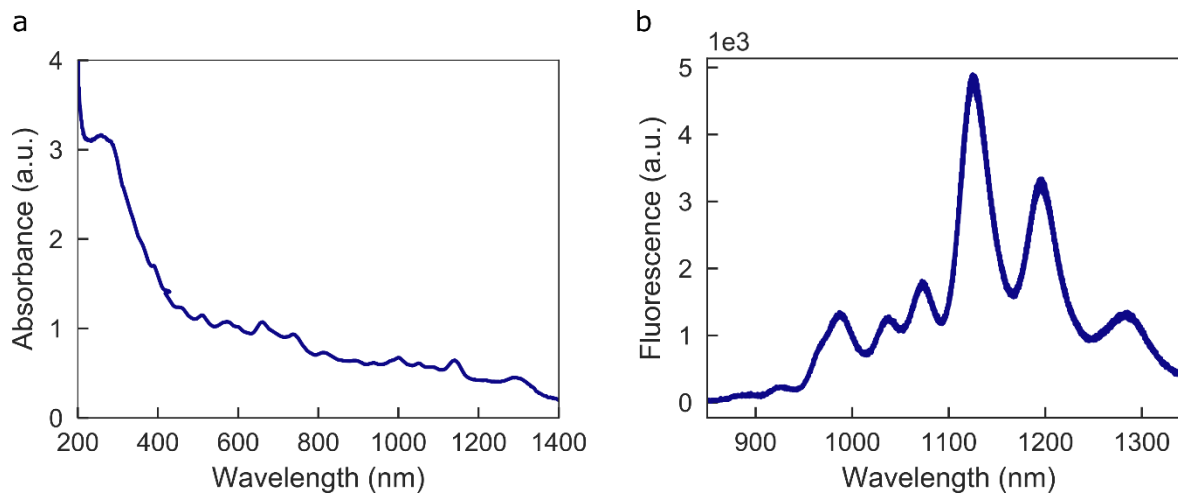
Figure 5-2. Optical characterization of $(GT)_{15}$-SWCNTs. (a) Absorbance spectrum of 30 mg L^{-1} $(GT)_{15}$-SWCNTs in PBS. (b) Fluorescence spectrum of 10 mg L^{-1} $(GT)_{15}$-SWCNTs in PBS. Stable SWCNT suspension in aqueous medium is confirmed by absorbance peaks across the visible and near-infrared range and fluorescence emission that would otherwise be quenched in a SWCNT- aggregated state.
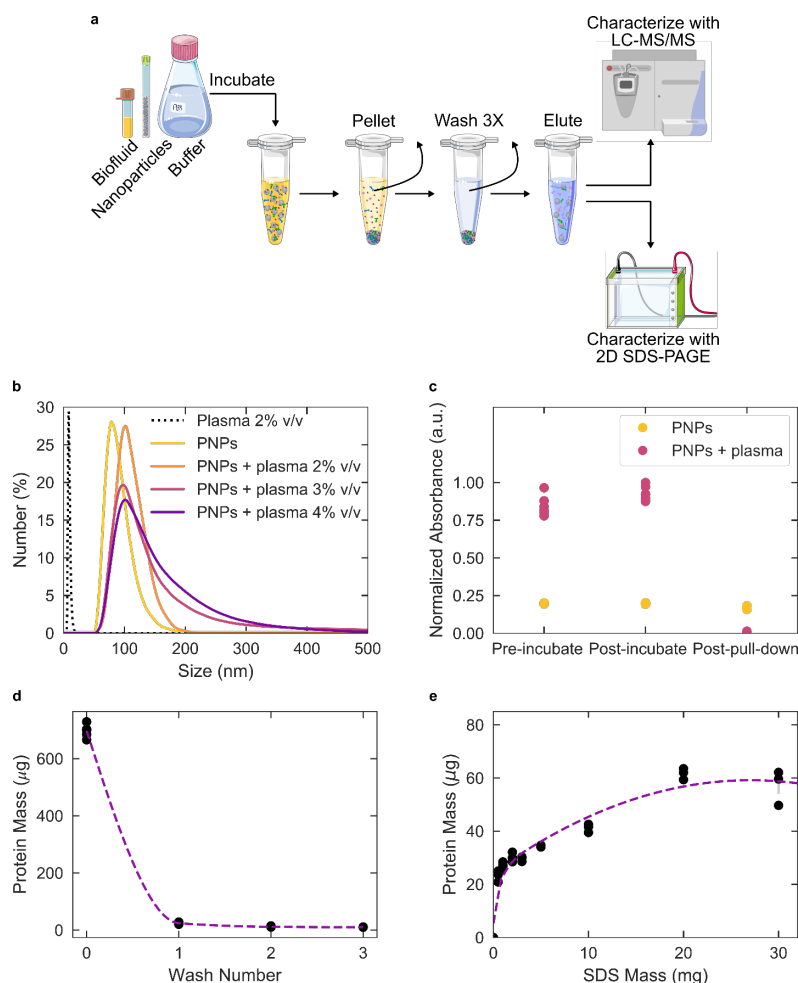
Figure 5-3. Isolation and characterization of protein-nanoparticle complexes to determine protein corona composition on nanoparticles. (a) Schematic detailing experimental procedure: nanoparticles are incubated with the desired biofluid in buffered solution, nanoparticle-protein complexes are pelleted by centrifugation and washed three times to remove non-selectively pelleted proteins, and corona proteins are eluted and characterized by two-dimensional polyacrylamide gel electrophoresis (2D PAGE) or liquid chromatography-tandem mass spectrometry (LC-MS/MS). (b) Dynamic light scattering (DLS) reveals that plasma protein corona formation induces an increase in the hydrodynamic radius of the PNPs (1.67 g L^{-1} in PBS) via peak shifting and broadening. (c) Absorbance at PNP excitation max (441 nm) immediately after adding plasma to incubation solution, incubating for 1 hour, and after the first pelleting step demonstrates the presence of proteins facilitates isolation of nanoparticles from solution in the initial pelleting step. (d) Quantification of free protein in solution via Qubit Protein Assay for varying wash number shows nearly complete depletion of free protein by three washes. (e) Quantification of eluted protein from nanoparticles via Pierce 660 nm Protein Assay with increasing SDS reducing buffer confirms complete elution of bound proteins from nanoparticle surface prior to characterization. Error bars on (b)-(d) are ± standard error for experimental replicates of N = 6, 6, and 3, respectively.

88
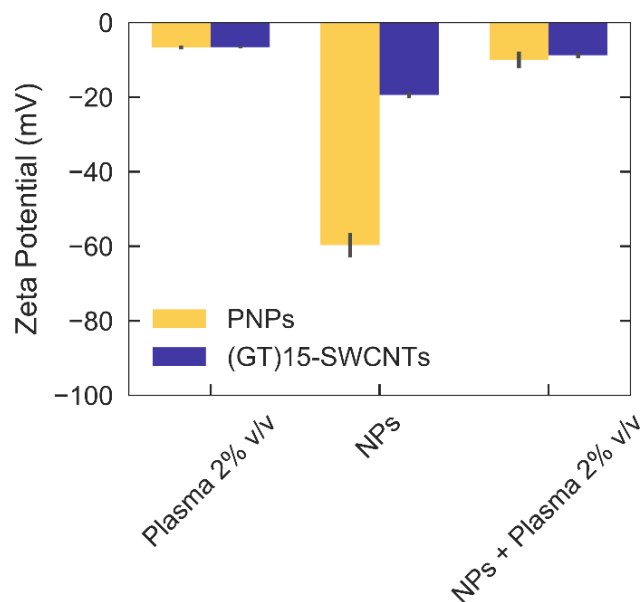
Figure 5-4. Surface charge changes induced by plasma protein corona formation. Zeta potential of native plasma, nanoparticles alone (PNPs yellow, $(GT)_{15}$-SWCNTs purple), and plasma protein-nanoparticle complexes. Lower magnitude zeta potential of protein-nanoparticle complexes indicates reduction in colloidal stability in the presence of surface-adsorbed proteins, as expected by visible aggregates formed. PNPs are 1.67 g L^{-1} and $(GT)_{15}$-SWCNTs are 28.67 mg L^{-1}, in PBS, 700 μL volume. Error bars are ± standard deviation for technical replicates (N = 3).

Table 5-4. Nanoparticle mass loss during corona isolation.

Wash	Mass [μg] PNPs (Plasma)	Mass [μg] PNPs (CSF)	Mass [μg] $(GT)_{15}$-SWCNTs (Plasma)	Mass [μg] $(GT)_{15}$-SWCNTs (CSF)
0	103.45 ± 29.40	37.51 ± 0.00	6.25 ± 0.21	3.33 ± 0.21
1	19.33 ± 1.97	0.00	0.00	0.00
2	11.37 ± 1.97	0.00	0.00	0.00
3	10.23 ± 0.00	0.00	0.00	0.00
Total Mass Removed	144.37	37.51	6.25	3.33
Initial Calculated Mass	1250	300	215	9.5
Estimated Mass Loss %	*11.55%*	*12.50%*	*29.07%*	*35.09%*
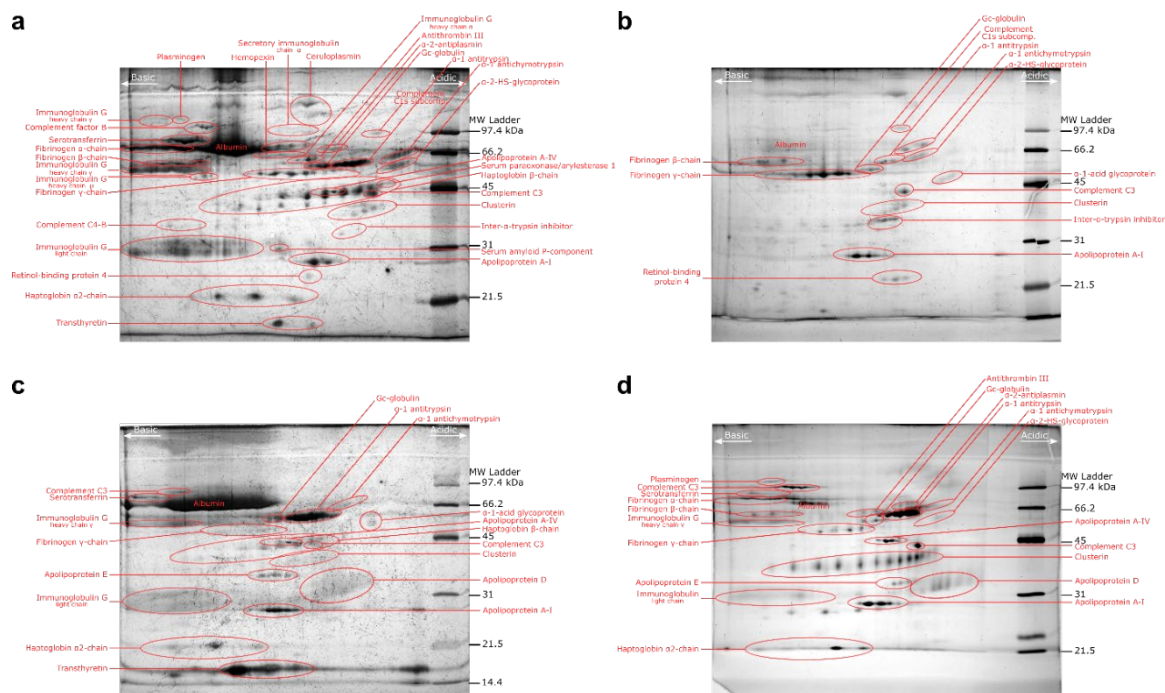
Figure 5-5. Representative 2D PAGE gels. (a) Plasma alone, (b) Plasma protein corona composition formed on (GT)$_{15}$-SWCNTs, (c) CSF alone, and (d) CSF protein corona composition formed on PNPs.

Table 5-5. Top 20 most abundant proteins identified by proteomic mass spectrometry in plasma $(GT)_{15}$-SWCNT and $(GT)_6$-SWCNT coronas.

	Plasma	$(GT)_{15}$-SWCNTs in plasma	$(GT)_6$-SWCNTs in plasma
1	Serum albumin	Clusterin	A disintegrin and metalloproteinase with
2	Haptoglobin	Histidine-rich glycoprotein	Apolipoprotein A-I
3	Ig kappa constant	Apolipoprotein A-I	Complement C3
4	Ig heavy constant gamma	Complement C3	Clusterin
5	Serotransferrin	Haptoglobin	Histidine-rich glycoprotein
6	Apolipoprotein A-I	A disintegrin and metalloproteinase with thrombospondin motifs 12	Prothrombin
7	Complement C4	Complement C1r subcomponent	Kininogen-1
8	Telomeric repeat-binding factor 2-interacting protein	Vitronectin	C4b-binding protein alpha chain
9	Alpha-1-antitrypsin	Kininogen-1	Vitronectin
10	Alpha-2-HS-glycoprotein	Prothrombin	Haptoglobin
11	Apolipoprotein A-II	C4b-binding protein alpha chain	Fibrinogen alpha chain
12	Ig heavy constant alpha 1	Complement factor H	Ig J chain
13	Integrin alpha-7	Fibrinogen alpha chain	Complement C1r subcomponent
14	Alpha-2-macroglobulin	Protein AMBP	Apolipoprotein E
15	Complement C3	Beta-2-glycoprotein 1	Beta-2-glycoprotein 1
16	Complement C5	Apolipoprotein E	Ig heavy constant gamma 1
17	Hemopexin	Complement C1q subcomponent subunit B	Alpha-2-HS-glycoprotein
18	Alpha-1-acid glycoprotein 1	Ig heavy constant gamma 1	Transthyretin
19	Ig heavy constant mu	Ig J chain	Protein AMBP
20	Beta-2-glycoprotein 1	Galectin-3-binding protein	Alpha-1-acid glycoprotein 2
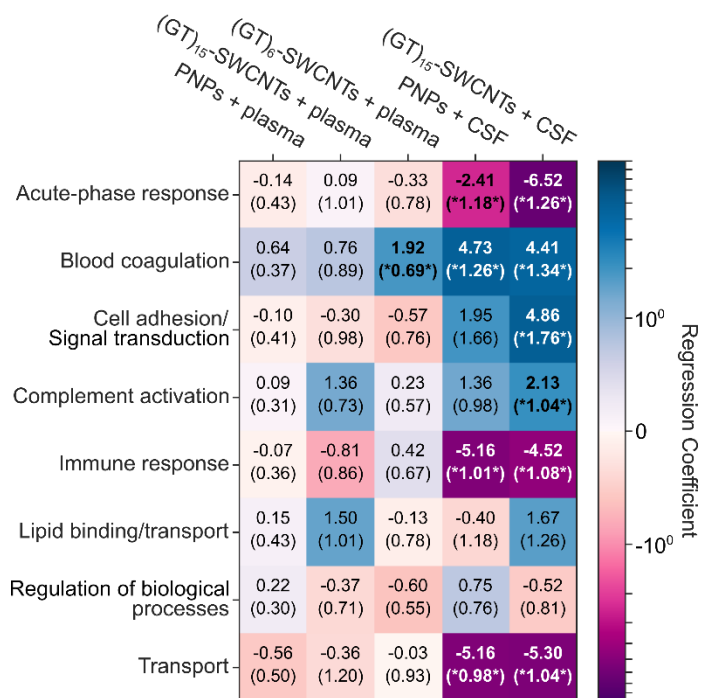
Figure 5-6. Role of protein functional class in protein corona formation for each nanoparticle-biofluid pairing. Ln-fold change, effect-coded regression coefficients of protein classes (rows) for each nanoparticle-biofluid pairing (columns). Cells are colored from dark purple (lower than the average fold change) to white (average fold change) to dark blue (higher than average fold change). Standard errors of the coefficients are given in parentheses. Results that have false-discovery-rate-corrected p-values of below 0.1 are bolded and noted with asterisks.
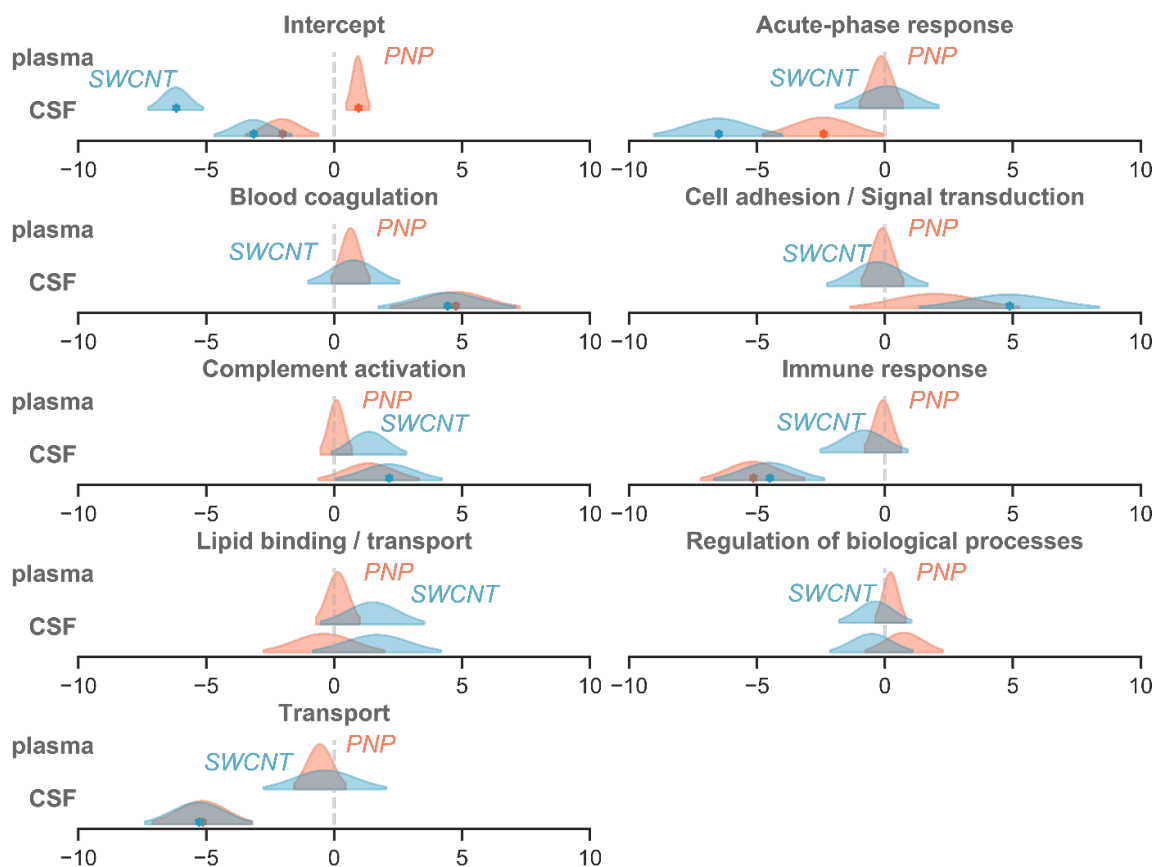
Figure 5-7. Distribution for protein class mean regression coefficients in each nanoparticle-biofluid pairing. Stars indicate false-discovery-rate adjusted p-values<0.1.
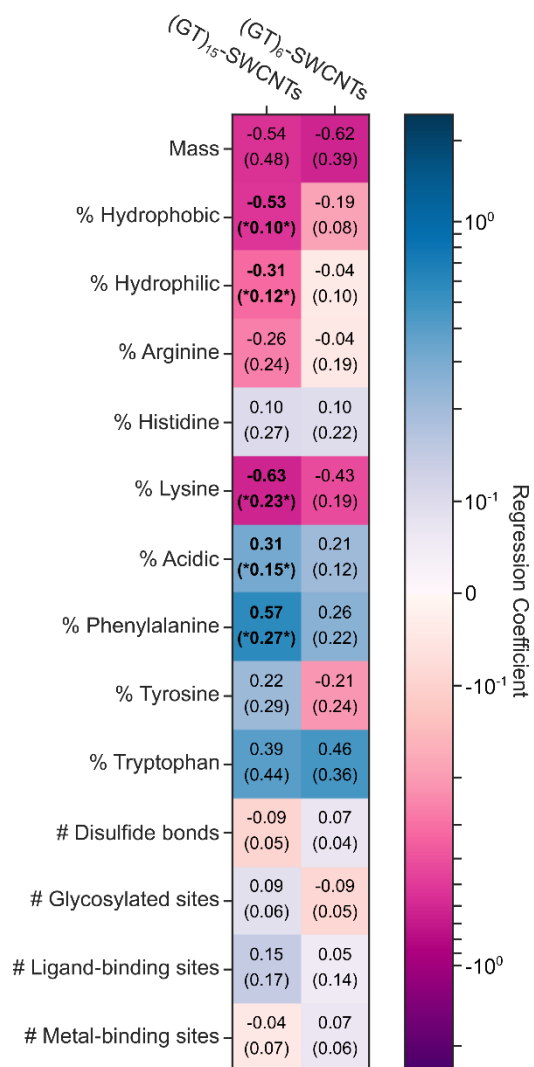
Figure 5-8. Molecular attributes of proteins that govern protein corona formation for $(GT)_x$-SWCNTs in plasma. Ln-fold change regression coefficients for molecular attributes of proteins (rows) for each nanoparticle-biofluid pairing (columns). Cells are colored from dark purple (negative effect on fold change) to white (no effect) to dark blue (positive effect). Standard errors of the coefficients are given in parenthesis. Results that have false-discovery-rate-corrected p-values below 0.1 are bolded and noted with asterisks. Amino acid groupings include: non-aromatic hydrophobic (sum of alanine, valine, isoleucine, leucine, and methionine content), hydrophilic (sum of serine, threonine, asparagine, glutamine content), and acidic (sum of aspartic acid and glutamic acid content).
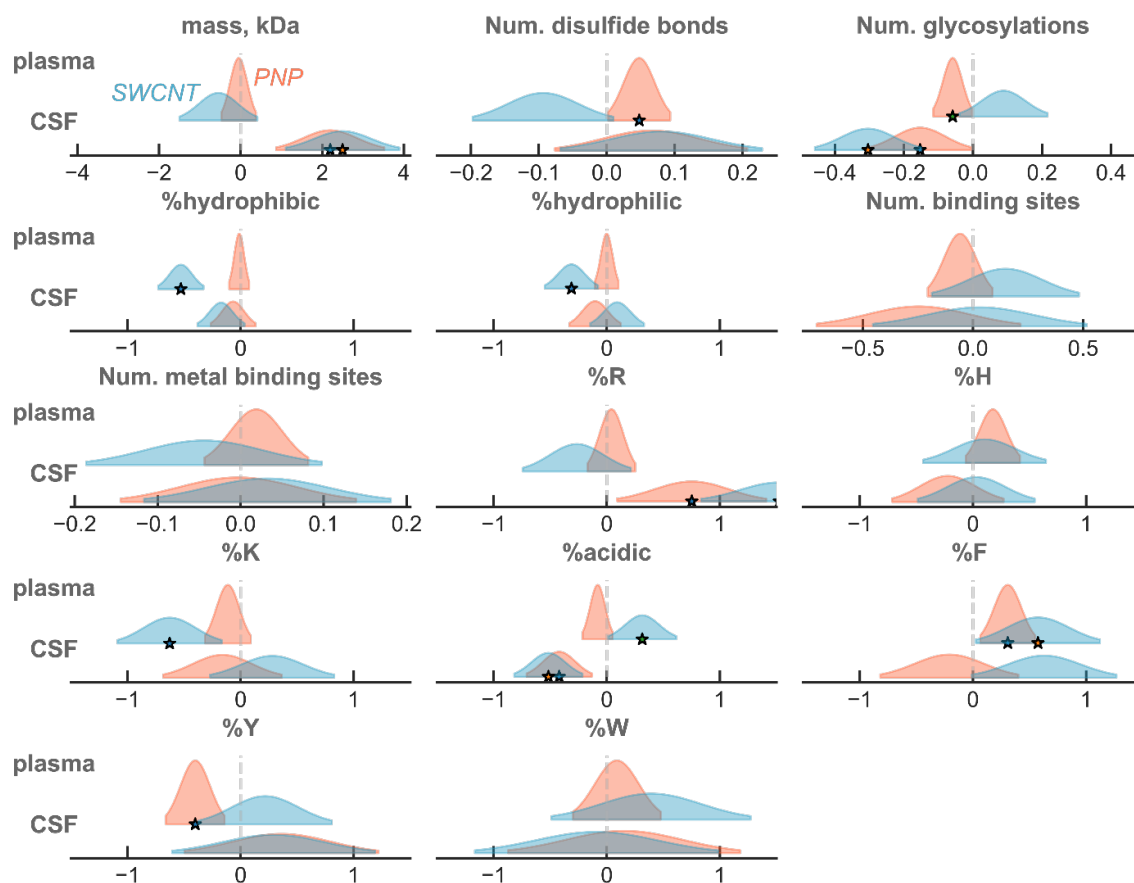
Figure 5-9. Distribution for microscale mean regression coefficients in each nanoparticle-biofluid pairing. Stars indicate false-discovery-rate adjusted p-values<0.1.

Table 5-6. Protein class regression results for each nanoparticle-biofluid pairing.

	PNPs in plasma			$(GT)_{15}$-SWCNTs in plasma		
	R-squared	Adjusted R-squared		R-squared	Adjusted R-squared	
	0.26	0.24		0.1	0.13	
	Parameter	Standard Error	FDRC p-values	Parameter	Standard Error	FDRC p-values
Intercept	0.9277	0.2270	0.0003	-6.1942	0.5380	0.0000
Sample 1	2.8743	0.3253	0.0000	4.0556	0.7711	0.0000
Sample 2	-0.0358	0.3146	0.9095	0.3037	0.7457	0.8412
Acute-phase response	-0.1393	0.4271	0.9095	0.0916	1.0124	0.9279
Blood coagulation	0.6394	0.3749	0.3267	0.7639	0.8887	0.7163
Cell adhesion / Signal transduction	-0.0981	0.4133	0.9095	-0.3003	0.9797	0.8412
Complement activation	0.0854	0.3086	0.9095	1.3564	0.7316	0.2372
Immune response	-0.0727	0.3620	0.9095	-0.8149	0.8580	0.7163
Lipid binding / transport	0.1467	0.4271	0.9095	1.5017	1.0124	0.3823
Regulation of biological processes	0.2203	0.2985	0.9095	-0.3749	0.7075	0.8412
Transport	-0.5563	0.5047	0.7460	-0.3584	1.1965	0.8412
	PNPs in CSF			$(GT)_{15}$-SWCNTs in CSF		
	R-squared	Adjusted R-squared		R-squared	Adjusted R-squared	
	0.32	0.28		0.35	0.31	
	Parameter	Standard Error	FDRC p-values	Parameter	Standard Error	FDRC p-values
Intercept	-2.0347	0.7108	0.0131	-3.1619	0.7551	0.0001
Sample 1	0.9624	0.8879	0.3850	-0.0962	0.9431	0.9939
Sample 2	0.8264	0.8915	0.3909	-0.0073	0.9470	0.9939
Acute-phase response	-2.4143	1.1830	0.0943	-6.5197	1.2566	0.0000
Blood coagulation	4.7294	1.2588	0.0009	4.4095	1.3371	0.0026
Cell adhesion / Signal transduction	1.9484	1.6552	0.3785	4.8610	1.7583	0.0117
Complement activation	1.3603	0.9806	0.3067	2.1252	1.0417	0.0675
Immune response	-5.1599	1.0141	0.0000	-4.5191	1.0772	0.0001
Lipid binding / transport	-0.3967	1.1830	0.7378	1.6716	1.2566	0.2548
Regulation of biological processes	0.7459	0.7590	0.3909	-0.5229	0.8062	0.6325
Transport	-5.1597	0.9806	0.0000	-5.3045	1.0417	0.0000

Table 5-7. Microscale regression results for each nanoparticle-biofluid pairing.

	PNPs in plasma			$(GT)_{15}$-SWCNTs in plasma		
	R-squared	Adjusted R-squared		R-squared	Adjusted R-squared	
	0.32	0.29		0.27	0.23	
	Parameter	Standard Error	FDRC p-values	Parameter	Standard Error	FDRC p-values
Intercept	2.7690	3.6300	0.7585	19.6116	8.2339	0.0607
Sample 1	2.8759	0.3147	0.0000	4.0662	0.7137	0.0000
Sample 2	-0.0427	0.3036	0.9436	0.3186	0.6886	0.6842
Mass	-0.0475	0.2109	0.9314	-0.5446	0.4785	0.4153
% hydrophobic residues (nonaromatic)	-0.0126	0.0442	0.9314	-0.5286	0.1003	0.0000
% hydrophilic residues	-0.0008	0.0520	0.9877	-0.3122	0.1179	0.0363
% arginine	0.0409	0.1051	0.9121	-0.2643	0.2385	0.4153
% histidine	0.1762	0.1197	0.4030	0.1028	0.2716	0.7053
% lysine	-0.1128	0.1015	0.5678	-0.6276	0.2302	0.0363
% acidic residues	-0.0806	0.0662	0.5447	0.3131	0.1502	0.0921
% phenylalanine	0.3077	0.1209	0.0648	0.5747	0.2743	0.0921
% tyrosine	-0.4013	0.1297	0.0184	0.2165	0.2942	0.5615
% tryptophan	0.0893	0.1940	0.9121	0.3912	0.4400	0.4942
Number of disulfide bonds	0.0477	0.0229	0.1288	-0.0938	0.0519	0.1527
Number of glycosylated sites	-0.0588	0.0280	0.1288	0.0885	0.0636	0.3124
Number of ligand binding sites	-0.0584	0.0736	0.7585	0.1473	0.1668	0.4942
Number of metal binding sites	0.0190	0.0313	0.8425	-0.0442	0.0710	0.6052

	PNPs in CSF			$(GT)_{15}$-SWCNTs in CSF		
	R-squared	Adjusted R-squared		R-squared	Adjusted R-squared	
	0.35	0.29		0.4	0.34	
	Parameter	Standard Error	FDRC p-values	Parameter	Standard Error	FDRC p-values
Intercept	-18.8198	9.7672	0.1897	-32.4292	10.2373	0.0063
Sample 1	0.7932	0.8845	0.5921	-0.4235	0.9271	0.8692
Sample 2	0.7215	0.8883	0.5921	-0.2622	0.9311	0.9097
Mass	2.1956	0.6640	0.0199	2.4973	0.6960	0.0025
% hydrophobic residues (nonaromatic)	-0.0670	0.0987	0.6048	-0.1726	0.1034	0.2361
% hydrophilic residues	-0.1034	0.1133	0.5921	0.0913	0.1187	0.7529
% arginine	0.7511	0.3295	0.1358	1.5253	0.3453	0.0003
% histidine	-0.2217	0.2464	0.5921	0.0293	0.2583	0.9097
% lysine	-0.1607	0.2616	0.6118	0.2798	0.2742	0.5836
% acidic residues	-0.4209	0.1441	0.0341	-0.5162	0.1511	0.0034
% phenylalanine	-0.2088	0.3052	0.6048	0.6294	0.3199	0.1441
% tyrosine	0.3624	0.4282	0.5921	0.2910	0.4488	0.8001
% tryptophan	0.1523	0.5128	0.8149	-0.0914	0.5375	0.9097
Number of disulfide bonds	0.0653	0.0708	0.5921	0.0801	0.0742	0.5836
Number of glycosylated sites	-0.1529	0.0733	0.1644	-0.3039	0.0769	0.0010
Number of ligand binding sites	-0.2462	0.2313	0.5921	0.0309	0.2424	0.9097
Number of metal binding sites	-0.0028	0.0709	0.9685	0.0323	0.0744	0.8692
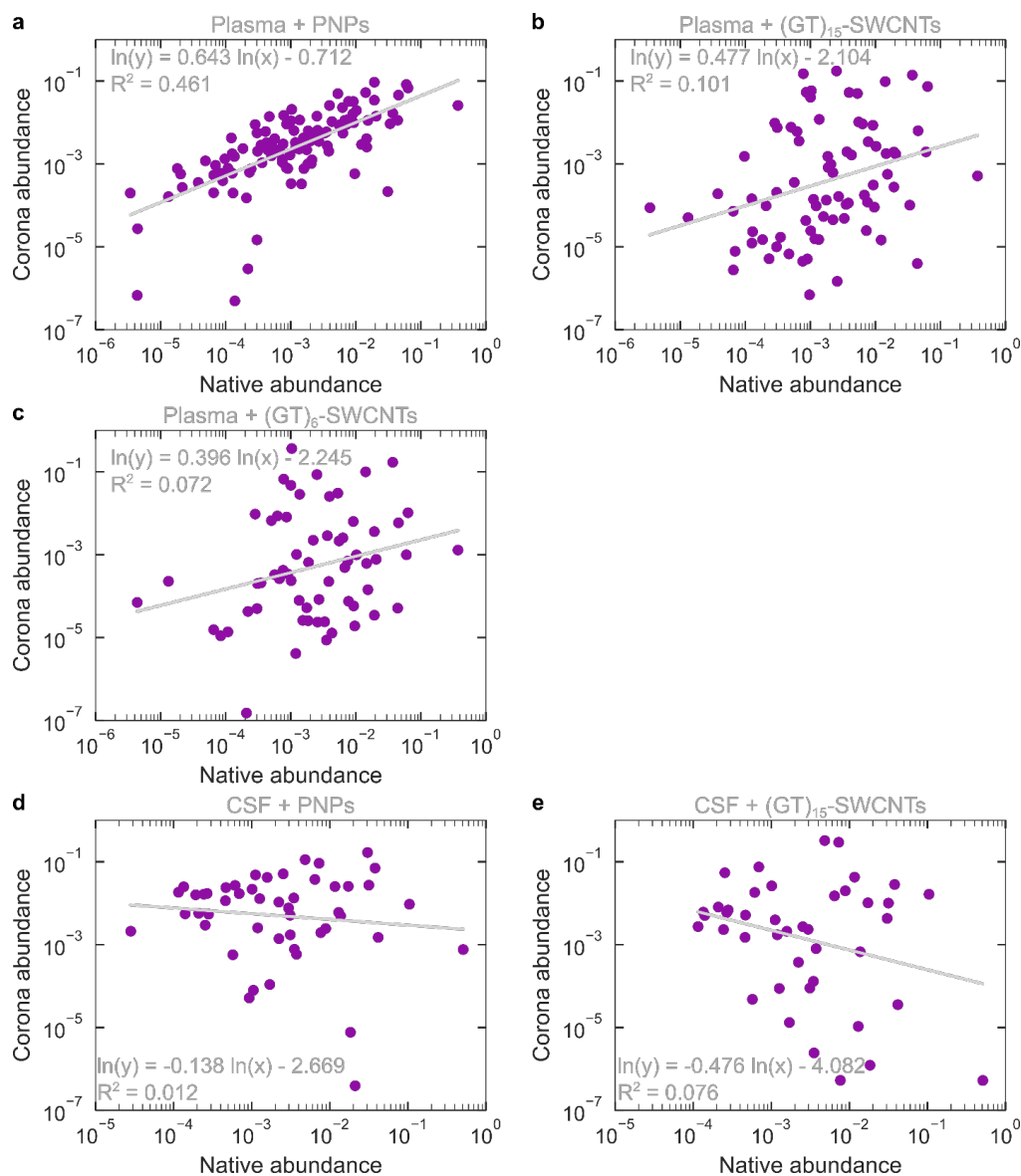
Figure 5-10. Scaling of protein abundance in corona vs. in native biofluid. Protein mole fraction of plasma proteins in corona of (a) PNPs, (b) (GT)15-SWCNTs, and (c) (GT)6-SWCNTs, vs. protein mole fraction of plasma proteins in native biofluid. Corona abundance scaling is approximately linear for plasma proteins on PNPs (R2 = 0.461) vs. highly scattered for (GT)15-SWCNTs (R2 = 0.101) and (GT)6-SWCNTs (R2 = 0.072). Protein mole fraction of CSF proteins in corona of (d) PNPs and (e) (GT)15-SWCNTs vs. protein mole fraction of CSF proteins in native biofluid. Corona abundance displays a weak negative correlation with native abundance for CSF proteins on both PNPs (R2 = 0.012) and (GT)15-SWCNTs (R2 = 0.076). All mole fractions are on a solvent-free basis. Note that proteins with zero corona abundance are excluded from the analysis for clarity, but the same conclusions hold when included.
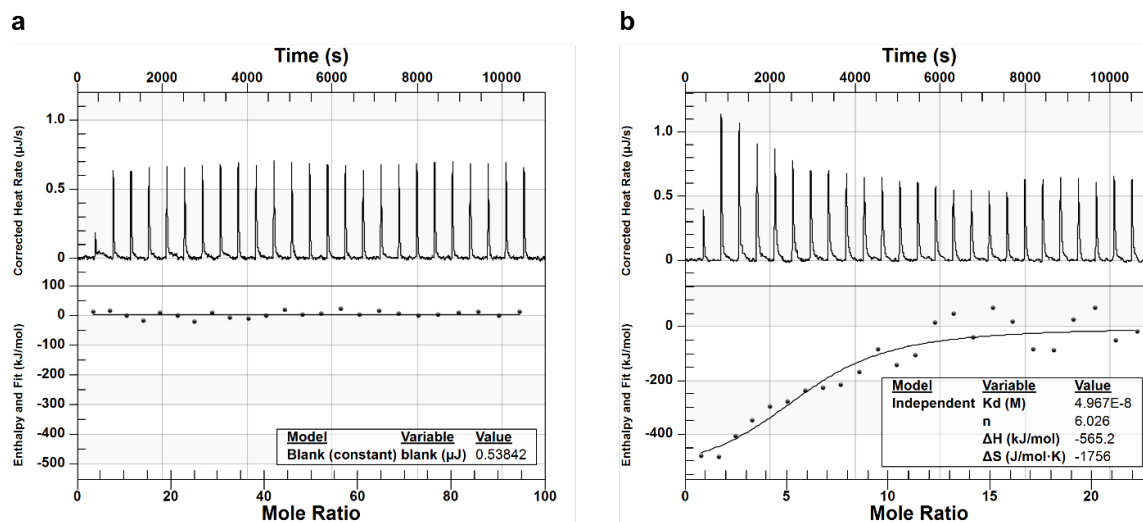
Figure 5-11. Protein corona thermodynamics assessed with ITC for binding of key proteins to (GT)$_{15}$-SWCNTs. Isothermal titration calorimetry (ITC) is employed to determine binding thermodynamics of (a) albumin and (b) fibrinogen to (GT)$_{15}$-SWCNTs. Albumin does not bind to (GT)$_{15}$-SWCNTs within experimentally accessible limits of this instrument, whereas fibrinogen does, in agreement with the corona compositional analyses from proteomic mass spectrometry and gel electrophoresis.
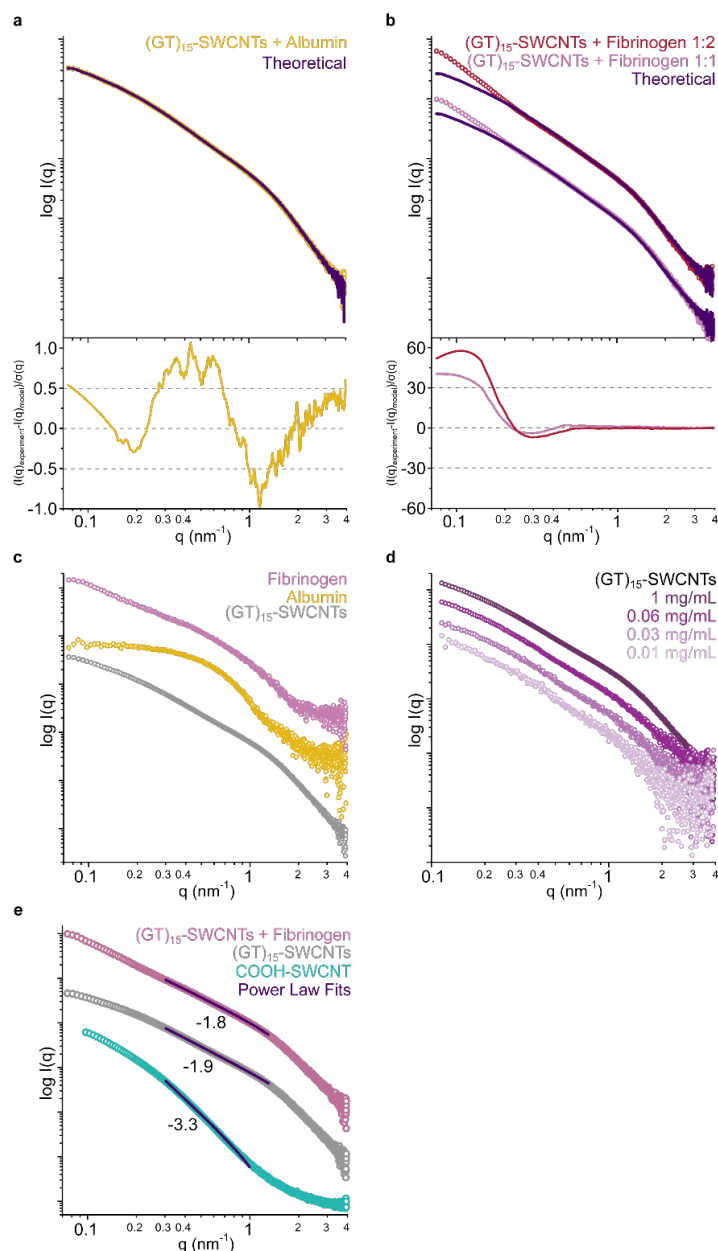
Figure 5-12. Protein corona structure assessed with SAXS for binding of key proteins to $(GT)_{15}$-SWCNTs. The linear combination of respective standard curves from panel c in purple and fit-residuals below, fit against the curves produced by the potential complexes of $(GT)_{15}$-SWCNTs with (a) albumin or (b) fibrinogen, at two different ratios of $(GT)_{15}$-SWCNTs to fibrinogen (1:1 is 0.5 g L^{-1} final concentrations of $(GT)_{15}$-SWCNTs and fibrinogen; 1:2 is 0.25 g L^{-1} $(GT)_{15}$-SWCNTs and 0.5 g L^{-1} fibrinogen). (c) Experimental SAXS profiles for standards of albumin, fibrinogen, and $(GT)_{15}$-SWCNTs alone, at identical concentrations to the mixing experiments (all 0.5 g L^{-1}). (d) SAXS profiles for concentration series of $(GT)_{15}$-SWCNTs alone, 0.01 – 1 g L^{-1}. (e) SAXS profiles fit to show power law dependencies in the Porod regions, including the COOH-SWCNT control without surface-adsorbed ssDNA.

100

Table 5-8. SAXS mass fractal modeling parameters.

Sample	Radius (nm)	Fractal Dimension (Dm)	Cutoff Length (nm)
(GT)$_{15}$-SWCNTs + Fibrinogen	1.05 ± 0.003	1.77	103.34 ± 9.70
(GT)$_{15}$-SWCNTs + Albumin	1.05 ± 0.003	1.90	10.60 ± 0.05
(GT)$_{15}$-SWCNTs	1.01 ± 0.002	1.89	10.91 ± 0.04
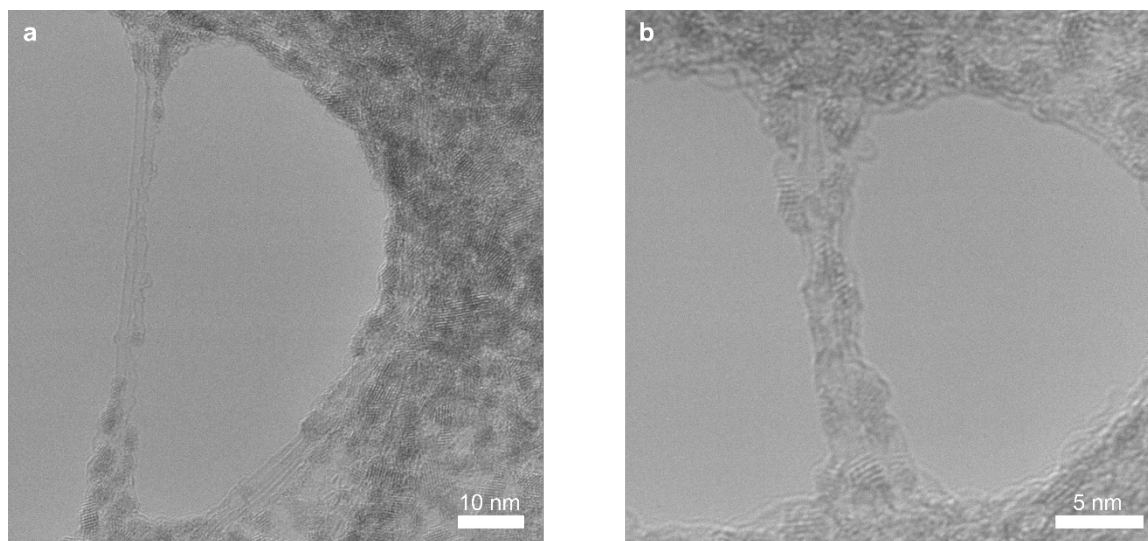
Figure 5-13. Protein corona morphology visualized by TEM for adsorption of plasma proteins to (GT)$_{15}$-SWCNTs. Transmission electron microscopy (TEM) of (a) plasma protein corona and (b) fibrinogen corona on (GT)$_{15}$-SWCNTs.

5.6.2   Extended Discussion on ITC

ITC was employed to extract relative binding parameters of protein-nanoparticle association. ITC was performed at constant pressure such that the heat absorbed or released is equivalent to the change in enthalpy (ΔH^o) upon binding. The binding curve can be fit to determine the equilibrium dissociation constant (K_d) and molar binding stoichiometry (n). This enables subsequent calculation of changes in standard state Gibbs free energy (ΔG^o) and entropy (ΔS^o) as follows:

$$\Delta G^o = \ RT\mathrm{ln}K_d = \Delta H^o - T\Delta S^o \qquad\qquad 5\text{-}3$$

where R is the ideal gas constant and T is temperature. The optimized run parameters to measure heats of binding for this system require relatively high protein and nanoparticle concentrations: for each run, 10 μL of 1.2 g L^{-1} protein was added for each of 24 injections from the syringe into 1 mL of 0.1 g L^{-1} (GT)$_{15}$-SWCNTs in the cell. At these concentrations, addition of fibrinogen causes visible sample aggregation, presumably due to polymer bridging interactions of proteins adsorbed on one nanoparticle interacting with another nanoparticle. One of the key assumptions of ITC is that the system is equilibrated during each titration step. Yet, aggregation is a kinetically controlled, non-equilibrium process. As the key assumption is not held, these binding values are actually the convolution of protein binding to individual SWCNTs, fibrinogen binding to aggregated SWCNTs, and SWCNTs aggregating. We can compensate for this limitation in data processing by applying the Lumry-Eyring model,[214] in which an equilibrium reaction is coupled to a self-association reaction (i.e. aggregation), and the heats measured are separated out accordingly. This encompasses subtracting out baseline aggregation heats and arriving at an apparent binding heat. Therefore, the thermodynamic parameters are reported with consideration of these higher order processes taking place simultaneously. A further note is that baseline drift/shift were observed during these ITC experiments involving (GT)$_{15}$-SWCNTs. These changes in baseline often indicate slow non-equilibrium processes in action, further confirming the presence of aggregation. In conclusion, ITC is not a suitable methodology to study nanoparticle-protein corona formation for all systems, and these limitations must be considered during experimental design and reporting of results.

5.6.3   Extended Experimental and Modeling Details and Discussion on SAXS

Experimental SAXS profiles were collected for 0.5 g L^{-1} (GT)$_{15}$-SWCNTs with and without albumin or fibrinogen, each at 0.5 g L^{-1} final concentrations (Figure 5-1c). The linear combination of (GT)$_{15}$-SWCNTs and albumin standard curves produced a SAXS profile identical to the mixed sample of (GT)$_{15}$-SWCNTs with albumin, suggesting no interaction between the species. Dissimilarly, no calculated linear combination of the (GT)$_{15}$-SWCNTs and fibrinogen standard curves could be produced to fit the SAXS profiles of the mixed sample, indicating formation of unique form factors and thus complexation. Additionally, a clear concentration dependence is observed with an increase in the ratio of fibrinogen to (GT)$_{15}$-SWCNTs by two-fold, while albumin shows no additional binding at elevated concentrations (Figure 5-12a-b). Control SAXS profiles of albumin, fibrinogen, and (GT)$_{15}$-SWCNTs alone were collected at identical concentrations to those of the mixing experiments (Figure 5-12c). Data was collected at elevated concentrations (0.5 g L^{-1} both protein and (GT)$_{15}$-SWCNTs) to

enhance SAXS signal, however, a concentration series was also performed for $(GT)_{15}$-SWCNTs to ensure that the scattering profiles do not deviate under more relevant nanoparticle conditions down to 0.01 g L^{-1} (Figure 5-12d).[166]

All $(GT)_{15}$-SWCNT samples with and without proteins were determined to be intrinsically disordered and experimental SAXS profiles were accordingly fit using mass fractal geometries. These fits were complemented by calculating power-law dependencies from the Porod region and were both calculated using the SasView software package (www.sasview.org). Scattering intensity as a function of scattering vector $I(q)$ calculations for the mass fractal modeling (Figure 5-1) was done as follows:[186]

$$I(q) = \text{scale} * P(q)S(q) + \text{background} \qquad\qquad 5\text{-}4$$

$$P(q) = F(qR)^2 \qquad\qquad 5\text{-}5$$

$$F(\text{x}) = \frac{3[\sin(x) - x\cos(x)]}{x^3} \qquad\qquad 5\text{-}6$$

$$S(q) = \frac{\Gamma(Dm - 1)\zeta^{Dm-1}}{[1 + (q\zeta)2]^{\frac{Dm-1}{2}}} \frac{\sin[(Dm - 1)\tan^{-1}(q\zeta)]}{q} \qquad\qquad 5\text{-}7$$

$$\text{scale} = \text{scale factor} * N\left(\frac{4}{3}\pi R^3\right)^2 (\rho_{\text{particle}} - \rho_{\text{solvent}})^2 \qquad\qquad 5\text{-}8$$

where R is the radius of the building block, Dm is the mass fractal dimension, ζ is the cut-off length, N is number of scatters, $\rho_{solvent}$ is the scattering length density of the solvent, and $\rho_{particle}$ is the scattering length density of particles. Dm relates the mass (m) to the radius as $m \sim R^{Dm}$ and is analogous to $I(q) \sim q^{-p}$ from the power-law calculations (with power-law exponent p), where $Dm = p$ when $q\zeta \gg 1$.

The power-law dependencies were determined by fitting the experimental SAXS profiles (Figure 5-12e), where $0.3 \leq q \leq 1$ nm^{-1} with the following:[187]

$$I(q) = \text{scale} * q^{-p} + \text{background} \qquad\qquad 5\text{-}9$$

These power-law dependencies (fits listed in Figure 5-12e) recapitulate the calculated Dm values from the mass fractal model fits.

Three main values are derived from these mass fractal and power-law calculations: (i) radius R (nm), (ii) fractal dimension Dm, and (iii) cutoff length ζ (nm) (Table 5-8).[186-188] The radius R in the mass fractal analysis is traditionally defined as the radius of the uniform sphere used to cover the fractal. The fractal dimension Dm and analogous power-law exponent p estimate the overall bulk geometries of the mass fractals, where the integer values of these variables represent the three dimensions in Euclidean space. Thus, Dm or $p = 1$, 2, or 3 represent rod, disk, or sphere

geometries, respectively. The cutoff length ζ defines the maximum distance between any two points of the mass fractal.

As another control, carboxylic acid functionalized SWCNTs (COOH-SWCNTs) were also examined via power-law scattering obtaining $p \sim 3.3$ (Figure 5-12e). This fit suggests that without ssDNA functionalization, COOH-SWCNTs form roughly spherical aggregates better modeled as a uniform density as opposed to a polymeric mass fractal. Thus, it may be inferred that ssDNA provides some semblance of order to the fine molecular structure of the system and should be the subject of further investigation.

The effect of aggregation on the scattering vector at very small angles ($q < 1$ nm^{-1}) precluded the use of the Guinier approximation and subsequent calculated metrics such as the radius of gyration (Rg), and the scattering intensity at $q=0$, $I(0)$, which is proportional to the molecular weight.[1] Additionally, while mathematically possible to calculate a pair-distribution function, $P(r)$, from the indirect Fourier transformation, the level of aggregation leads to non-zero values for $r=Dmax$.[215] Accordingly, we fit the whole SAXS profile to a specific mass fractal model, providing an estimate for the average cutoff length ζ, superseding the need to calculate the analogous $Dmax$ value which we determined to be less accurate.

6 Mapping the Morphology of DNA on Carbon Nanotube-Based Sensors in Solution using X-ray Scattering Interferometry **

6.1 Chapter Abstract

Single-walled carbon nanotubes (SWCNTs) with adsorbed single-stranded DNA (ssDNA) are applied as sensors to investigate biological systems, with applications ranging from clinical diagnostics to agricultural biotechnology. Unique ssDNA sequences render SWCNTs selectively responsive to target analytes. However, it remains unclear how the ssDNA conformation on the SWCNT surface contributes to their ultimate functionality, as observations have been constrained to computational models or experiments under dehydrated states that differ substantially from the aqueous biological environments in which the nanosensors are applied. Herein, we demonstrate a direct mode of measuring in-solution ssDNA geometries on SWCNTs via X-ray scattering interferometry (XSI), which leverages the interference pattern produced by AuNP tags conjugated to ssDNA on the SWCNT surface. We employ XSI to quantify distinct surface-adsorbed morphologies for two ssDNA oligomer lengths, conformational changes as a function of ionic strength, and the mechanism of dopamine sensing for a previously established ssDNA-SWCNT nanosensor, with corresponding *ab initio* modeling for visualization. We show that the shorter oligomer, $(GT)_6$, adopts a highly ordered structure of stacked rings along the SWCNT axis, compared to the longer, less periodic $(GT)_{15}$ wrapping. The presence of dopamine elicits a simultaneous axial elongation and radial constriction of the ssDNA closer to the SWCNT surface. Application of XSI to probe solution-phase morphologies of nanoparticle-based tools will yield insights into sensing mechanisms and inform future design strategies for polymer-functionalized SWCNT technologies.

6.2 Introduction

Single-walled carbon nanotubes (SWCNTs) serve as tools for biological sensing, imaging, and delivery applications.[199,216] SWCNTs are an advantageous platform due to their sensitive fluorescence response to localized changes (motivating sensor development[167,190,217]), photostable near-infrared fluorescence in the tissue-transparency window (enabling *in vivo* imaging[175,176]), and nanometer-sized diameter with a high aspect ratio (supporting use as cell-permeable delivery vehicles[218-220]). For each of these respective applications, the nanotube surface acts as a substrate upon which sensing moieties, anti-biofouling ligands, or delivery cargoes are loaded. Specifically, SWCNTs with adsorbed nucleic acids have been applied as nanoparticle-based sensors and delivery agents. Polymer properties including nucleic acid sequence and length govern SWCNT-adsorbed morphology, stability, and function. These constructs have proven particularly useful as nanosensors for small-molecule analytes including catecholamines[166,221-223], serotonin,[177,224] hydrogen peroxide,[225-227] and nitric oxide[170,228].

Despite over a decade of development in SWCNT-based sensors, there remain contrasting theories in the field about what enables molecular recognition, and what role (if any) conformational shifts play over chemical mechanisms. For example, a particular sequence of single-stranded DNA (ssDNA) – a repeating motif of guanine and thymine (GT) – has enabled highly sensitive and spatially resolved dopamine detection from single neurons and in neuronal tissue.[166,221-223] Hypothesized interaction mechanisms between the dopamine and GT oligomer include dual hydrogen bonding between the two hydroxyl groups of dopamine and the phosphate backbone of the ssDNA,[221] a redox reaction,[167] and/or intercalation of the aromatic catecholamine ring between the ssDNA oligomer and SWCNT surface driven by π-π stacking.[167] Optimizing interactions of nucleic acids with SWCNTs is key to the success of these biotechnologies, yet challenges remain in directly measuring, in real time, how ssDNA-SWCNT sensors behave.

Current methods for characterizing ssDNA-SWCNT conformations involve a dehydrated sample immobilized on a two-dimensional substrate, despite SWCNT-based biotechnologies mainly being applied in the aqueous solution state. Such techniques include transmission electron microscopy (TEM) to visualize ssDNA-SWCNT morphology[195,229] and atomic force microscopy (AFM) to determine dimensions and packing of biomolecules on SWCNTs,[220,230-232] which has been previously demonstrated to be limited by adsorption biases introduced during sample preparation.[194] Other physical properties such as hydrodynamic dimensions can potentially be extracted from dynamic light scattering (DLS) measurements on SWCNTs done in the solution state, however, rigorous optical scattering methods have not been well-adapted for non-spherical, high-aspect-ratio particles such as SWCNTs and cannot resolve fine-grained surface features such as nanometer-scale polymer packing.

Small-angle X-ray scattering (SAXS) has shown promise in revealing the morphology of SWCNT-based systems in solution.[233,234] We have previously reported the use of SAXS to determine the in-solution structure of ssDNA-suspended SWCNTs interacting with blood plasma proteins (see Chapter 5).[235] However, SAXS is a contrast measurement technique relying on the scattering intensity of the analyte (proportional to the square of the electron density) being significantly higher than that of the solution. Thus, materials of relatively low electron density such as carbon-based SWCNTs and ssDNA must be at sufficiently high concentrations for the signal to be above background. Accordingly, characterizing ssDNA-SWCNTs via SAXS require the use of SWCNT concentrations that exceed those actually applied in biological systems (0.1-5 mg/L).[166,216,236] These elevated concentrations can lead to artifacts such as inter-tube bundling of the ssDNA-SWCNTs,[229,233] which must be minimized to fully elucidate the morphology of individual ssDNA-functionalized SWCNT sensors. A strategy to overcome this concentration issue for low-scattering materials is to increase the X-ray exposure time, but this risks creating chemical changes in solution that can affect the sample under study.[1] An alternative approach is

to increase the electron density of the sample directly using high contrast materials such as gold nanoparticles (AuNPs) and then apply X-ray scattering interferometry (XSI), originally described by Mathew-Fenn *et al.*[29,30] XSI leverages the interference patterns generated upon X-ray scattering between ordered AuNPs to measure discrete inter-AuNP distances, effectively turning the AuNPs into molecular rulers in solution.[12,29,30,39,82,83,136] Additionally, through adaptation of robotics and a rapid data-processing pipeline, XSI can be run at higher throughput with minimal sample consumption (~1 sample per min with 25 µL per sample).[10,12,43]

Herein, we apply XSI to investigate nanopatterning of the adsorbed ssDNA corona surrounding the SWCNT surface in the solution phase. Small AuNPs (5.9-7.2 nm diameter) are attached to the 5' end of each ssDNA oligomer[12,136] and the ssDNA is adsorbed to the SWCNT surface, forming an ssDNA-SWCNT suspension with one AuNP tag per ssDNA strand. We focus on an illustrative example of how surface-constrained polymer conformation influences sensor properties by studying two $(GT)_n$ ssDNA sequences (n = 6, 15) used for dopamine sensing. These two ssDNA oligomers empirically possess different advantageous properties, with $(GT)_6$ displaying a larger magnitude of fluorescence change in response to dopamine[168] and $(GT)_{15}$ displaying higher stability in relevant biomolecule-rich environments.[178] Previous characterization by both experimental studies (AFM, EM)[195,231] and molecular dynamics (MD) simulations[168,237] has led to postulation that these two oligomers possess distinct surface-constrained conformations: the shorter $(GT)_6$ oligomer is expected to form a ring-like structure around the nanotube and the longer $(GT)_{15}$ oligomer is expected to form a helical wrapping around the nanotube. We employ high-throughput XSI in solution at biologically applicable concentrations to explore: (i) the configuration of adsorbed ssDNA along the SWCNT surface, (ii) the conformational changes of adsorbed ssDNA as a function of ionic strength, and (iii) the behavior of adsorbed ssDNA in the presence of the target analyte, dopamine. Additionally, we perform *ab initio* modeling of the AuNPs adsorbed on the SWCNT surface directly from scattering profiles to provide a more comprehensive, 3D view of the system. We validate our technique with other suspension characterization (absorbance, fluorescence, DLS) and direct visualization (TEM). Taken together, this approach establishes a high-throughput technique for in-solution characterization of these nanoparticle-based biotechnologies and provides a deeper understanding surrounding the mechanisms behind their molecular recognition.

6.3   Results and Discussion

6.3.1   Synthesis and characterization of ssDNA-AuNP-SWCNTs

To first demonstrate the resolvable concentration range for $(GT)_{15}$ ssDNA and $(GT)_{15}$-SWCNTs without AuNP tags, we collected SAXS profiles of serial dilutions for each (Figure 6-4A-B). Bundled ssDNA-SWCNTs were observed even at the lowest resolvable concentration of 16 mg/L, with average bundling of ~6-8 SWCNTs obtained from the cross-sectional radius of gyration (see SI Methods Section 6.5). No scattering contribution from the ssDNA alone was detected at equivalent concentrations (4 µM). These results and electron density calculations (see SI Methods Section 6.5 and Extended Discussion Section 6.6) motivate our use of small AuNP tags (5.9-7.2 nm diameter) to increase the electron density of our material and thus observe

ssDNA-SWCNTs in solution at relevant applied concentrations (<5 mg/L SWCNTs). Absolute-scale intensity scattering measurements demonstrate that scattering from 6.9 nm diameter PEGylated AuNPs (PEG-AuNPs) is ~270-fold higher than that of $(GT)_{15}$-SWCNTs and ~65,000-fold higher than that of $(GT)_{15}$ ssDNA (Figure 6-4C) at the correct relative concentrations (250 nM AuNP and ssDNA per 1 mg/L SWCNT; see SI Methods Section 6.5). Citrate-capped AuNPs were synthesized, conjugated to ssDNA via trithiolated linkers (Letsinger's type) on the 5' end, and coated with methoxy polyethylene glycol thiol (mPEG-SH) as detailed in SI Methods Section 6.5. Prepared ssDNA-AuNPs are then purified by anion exchange chromatography (Figure 6-5), characterized by SAXS to determine morphology and polydispersity (Figure 6-6 and Table 6-1), and DLS to confirm PEGylation by hydrodynamic radius (Figure 6-6 and Table 6-2) as detailed in SI Methods Section 6.5 and Extended Discussion Section 6.6. As anticipated, the measured scattering in the conjugated systems of ssDNA-AuNP-SWCNTs is dominated by the AuNP signal and eliminates the need to mathematically factor in the scattering contributions from the ssDNA or SWCNT alone, or the scattering cross-terms between the different components of the complex.

We apply XSI to study two ssDNA sequences based on their relevance to biomolecular sensing and predicted surface-adsorbed conformational differences: $(GT)_{15}$ and $(GT)_6$.[166,168,195,221] SWCNTs were suspended with ssDNA-AuNPs by probe-tip sonication at a constant ssDNA:SWCNT ratio (250 nmol ssDNA-AuNP per 1 mg SWCNT), in line with previous literature.[166,238] We optimized this suspension method with the added AuNP tag, as detailed in the SI Methods Section 6.5 and Extended Discussion Section 6.6. The resulting suspensions were characterized by absorbance and fluorescence analyses to corroborate ssDNA-AuNP-SWCNT complex formation (Figure 6-7). Retention of the AuNP plasmon resonance peak at approximately 520 nm reveals that the AuNP tags remain intact and well-dispersed through the SWCNT complexation process (Figure 6-7A). The apparent absence of absorbance peaks associated with SWCNT excitation is due to the high ratio of AuNPs to SWCNT in the ssDNA-AuNP-SWCNT complex and the limited dynamic range of the UV-Vis detector. Fluorescence spectra for AuNPs alone (with or without ssDNA) at 721 nm laser excitation reveal a trough in the emission intensity centered at approximately 950 nm. This optical feature may be due to absorption of excitation light, despite the lack of a distinct absorption band at this location. Compared to ssDNA-SWCNTs alone, addition of the AuNP tags results in lower intensity, broadened SWCNT fluorescence emission peaks, more prominently for $(GT)_6$- than $(GT)_{15}$-AuNP-SWCNTs and for emission peaks at shorter wavelengths (Figure 6-7B-D). This quenching effect of the AuNPs on the SWCNT fluorescence indicates electronic or excitonic interaction, and underscores the proximity of AuNPs to the SWCNT surface: AuNPs are metallic with known ultra-efficient quenching properties within 1-10s of nanometer-scale separation distances[239] and SWCNTs are sensitive to perturbations in their local dielectric environment.[240] The quenching mechanism may implicate photo-induced electron transfer or field effects of the proximal AuNPs on SWCNT excitons biasing toward nonradiative decay pathways,[241] with potential vibrational contributions to peak broadening. The greater degree of quenching and peak-broadening observed for the $(GT)_6$-AuNP-SWCNTs in comparison to the $(GT)_{15}$-AuNP-SWCNTs likely arises from different ssDNA surface packing, where more $(GT)_6$ ssDNA strands are expected per SWCNT based on previous literature (see Chapter 5).[193,235] This difference between $(GT)_{15}$- and $(GT)_6$-AuNP surface packing is also confirmed by our TEM analysis as

0.139 vs. 0.185 AuNPs per nanometer length of SWCNT, respectively (see SI Methods Section 6.5). The relative enhancement of fluorescence at longer wavelength peaks suggests more large-diameter SWCNTS individually dispersed with the ssDNA-AuNPs.

6.3.2   Conformational geometries of ssDNA on SWCNTs from XSI

The ssDNA-AuNP-SWCNT complexes were next analyzed by XSI. This technique is an extension of traditional solution SAXS in which a radial average of X-rays scattering off the electron density of a sample is integrated and the contrast between the sample and buffer is used to produce a buffer-subtracted 1D curve in reciprocal space. The total scattering intensity is the summation of two terms: the form factor, arising from the overall particle size and morphology, and the structure factor, derived from interparticle interactions. In structural biology, SAXS sample conditions (e.g., concentration) are adjusted to experimentally remove the contributions of the structure factor to isolate the form factor. An inverse Fourier transform of the subtracted curves then produces pairwise distribution functions ($P(r)$; probability plot of all inter-electron distances), providing real-space information on the average shape of the electron density of individual macromolecules free from interparticle interaction.[242,243] Conversely, in XSI, the structure factor is of primary importance and can be isolated to represent the interference pattern of scattered X-rays arising from inter-AuNP interactions, indicating discrete distances between ordered AuNPs.[12,29,136] In our case, we retain the form factor from the ssDNA-AuNP-SWCNT curves to enable normalization and additional analyses by preserving the information from the individual AuNPs.

Inter-ssDNA spacing along SWCNTs was measured using the AuNP tags via XSI (Figure 6-1A-B and SI Methods Section 6.5). Real-space analysis of the scattering profiles produces $P(r)$ functions with two main peaks (Figure 6-8). The first peak represents the intra-AuNP distances between electrons within individual AuNPs, with the peak maximum being the average radius of the AuNPs. The absence of additional peaks in the $P(r)$ functions without SWCNTs indicates that there is no long-range order and that the ssDNA-AuNPs (or PEG-AuNPs) are free in solution. The intra-AuNP peak provides a reference for the AuNP size distribution in each sample and enables normalization between samples to account for slight fluctuations in concentration and X-ray beam intensity. For clarity, the intra-AuNP peak is omitted in main figures but is included in supplementary figures. The second broader peak in the $P(r)$ functions represents the inter-AuNP distances and is only observed in complexes containing periodic ordering of AuNPs. This peak reveals the distinct surface-adsorbed spacings of $(GT)_{15}$- and $(GT)_6$-AuNPs on the nanotube surface (Figure 6-1A-B). Importantly, ssDNA-AuNPs (without SWCNT substrates) are in a disordered state when free in solution and only enter a periodically ordered state when adsorbed to the SWCNT surface (Figure 6-1A-B, Figure 6-8, and Figure 6-9). TEM visualization recapitulates these findings in the dried state showing AuNPs ordered (Figure 6-10 and Figure 6-11) or free (Figure 6-12).

A series of controls was analyzed to confirm that the preparation of ssDNA-AuNP-SWCNT complexes leads to adsorption of ssDNA-AuNPs on the SWCNT surface rather than off-target aggregative process: ssDNA-SWCNTs (no AuNPs), free ssDNA-AuNPs (no SWCNTs), and carboxylated SWCNTs mixed with ssDNA-AuNPs (no probe-tip sonication, and thus no driving force for self-assembly) do not reveal any feature suggesting AuNP order (Figure 6-13A-B).

Likewise, no order is observed in SWCNTs attempted-to-be suspended with PEG-AuNPs (no ssDNA) (Figure 1-13C-D). Additionally, there is no ssDNA-AuNP-SWCNT concentration dependence over the range used in this study (0.17-1.76 mg/L; Figure 6-14), which is within the unbundled SWCNT regime based on previous literature.[229] Finally, there is no AuNP size dependence for the axial inter-AuNP distances over the range used in this study (5.9-7.2 nm diameter; Figure 6-15A-C). Therefore, at this ssDNA:SWCNT ratio (250 nmol ssDNA-AuNP per 1 mg SWCNT), the packing of ssDNA on the SWCNT surface is not affected by potential steric effects from the AuNPs.
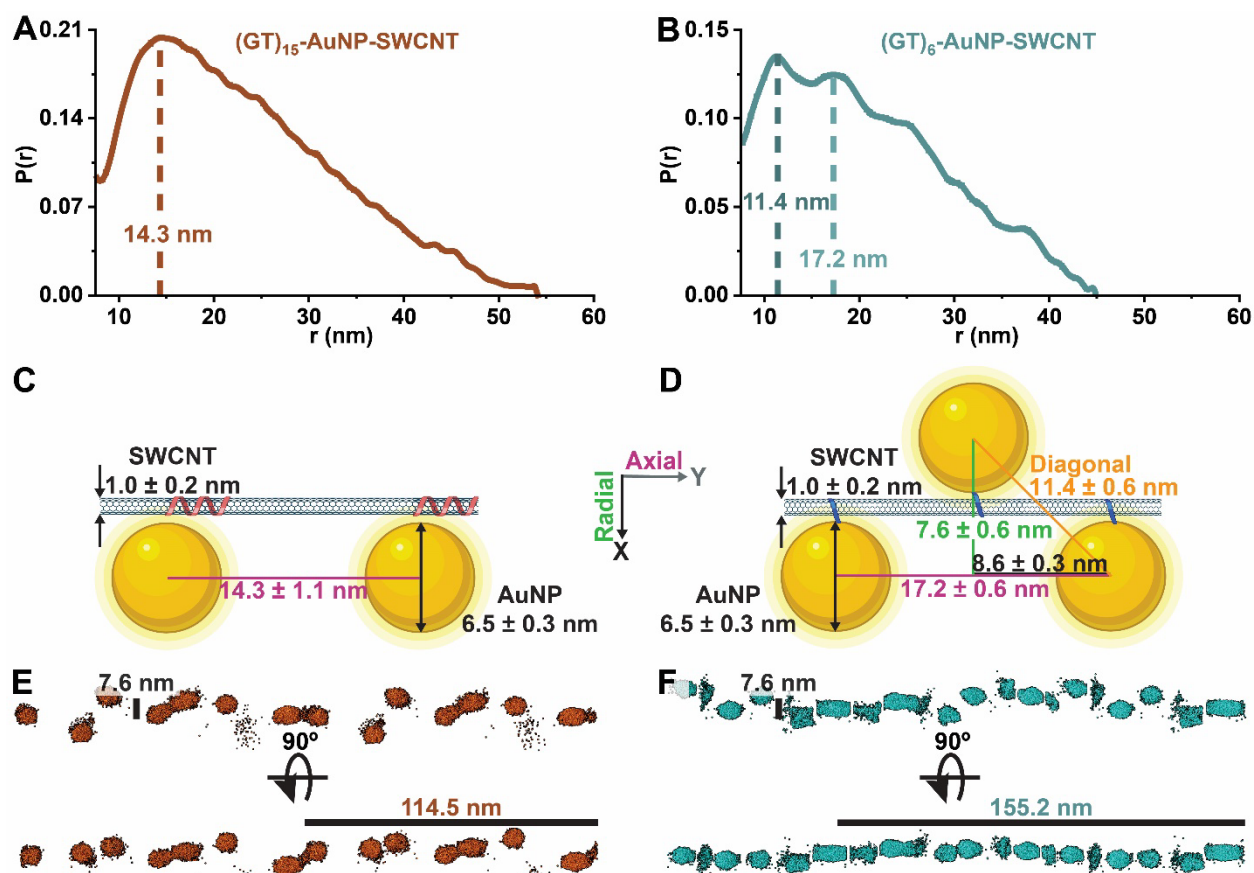
Figure 6-1. ssDNA forms ordered structures on the carbon nanotube surface. (A-B) Pairwise distribution functions, *P(r)*, from XSI data reveal discrete distances of AuNP-tagged ssDNA along the nanotube surface for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs. *P(r)* functions are normalized to the primary intra-AuNP peak, then the x-axis minimum is set to focus on the inter-AuNP peak for clarity. (C-D) 2D schematics for proposed geometrical arrangement of AuNPs on the SWCNT surface, with average inter-AuNP distances obtained from statistical analysis of *P(r)* functions for (C) (GT)$_{15}$-AuNP-SWCNT and (D) (GT)$_6$-AuNP-SWCNT. Average inter-AuNP distances are denoted as diagonal (orange), axial (magenta), and radial (green). Schematics are drawn to scale. (E-F) *Ab initio* modeling results for (E) (GT)$_{15}$-AuNP-SWCNT and (F) (GT)$_6$-AuNP-SWCNTs. Fits and residuals are shown in Figure 6-16.

110

MD simulations have shown that $(GT)_{15}$ forms a helical wrapping around SWCNTs with uniform electrostatic potential profiles, as opposed to the ring-like conformation of $(GT)_6$ showing a periodic electrostatic footprint.[168] For $(GT)_{15}$, this larger-footprint morphology and lower packing density (ascertained by TEM image analysis) is likely responsible for the generally broader and less defined inter-AuNP peaks observed for $(GT)_{15}$-AuNP-SWCNTs and suggests a more variable surface adsorption pattern (Figure 6-1A and Figure 6-10). Due to this increased variability in $(GT)_{15}$-AuNP adsorption and lack of orientational reference, a statistical analysis of the most probable inter-AuNP distances (14.3 ± 1.1 nm) is used to determine the basic 1D axial ssDNA spacing along the SWCNT (Figure 6-1C, Figure 6-15A, and C). In contrast, the inter-AuNP peaks for $(GT)_6$-AuNP-SWCNTs are narrower and contain more clearly defined higher-order features after the initial, most probable distance (11.4 ± 0.6 nm; Figure 6-1B and Figure 6-15B-C). MD simulations of $(GT)_6$ on (9,4) chirality SWCNTs (0.92 nm diameter) predict that there is a near-equivalent split in energetically favorable left-handed helix and ring-like conformations for low packing densities, and that steric effects from moderate-to-high surface coverage result in a population shift to primarily rings.[168] Accordingly, $(GT)_6$ is expected to adopt a ring-like configuration on the majority of SWCNTs used in this study (mixed chiralities, with average diameter of 1 ± 0.2 nm).

From the longer-range distance features of the inter-AuNP peak of $(GT)_6$-AuNP-SWCNTs, we deduce that the AuNPs align on alternating sides of the SWCNT. The second-most probable inter-AuNP distance reveals an average center-to-center distance of 17.2 ± 0.6 nm (Figure 6-1B and Figure 6-15B-C), representing the inter-AuNP spacing axially down the SWCNT and establishing the preceding peak at 11.4 nm the distance of AuNPs diagonally across the SWCNT (Figure 6-1D). These inter-AuNP distances are used to extrapolate the average periodic inter-ssDNA ring distance of 8.6 ± 0.3 nm (Figure 6-1D). As expected, the axial inter-AuNP distances for both $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs are not affected by changes in AuNP diameter (Figure 6-15C). Conversely, the diagonal distance is expected to change slightly as a function of the AuNP diameter, as calculated in Figure 6-15C when the AuNPs are flush with the SWCNT surface, holding to a 2D geometry (deemed adequate due to the large diameter disparity between AuNPs and SWCNTs). This trend is not seen experimentally, however, suggesting that the AuNPs are not directly in contact with the SWCNT surface, and their positional variance may preclude our ability to see this trend. To explore these dynamics, diagonal and axial inter-AuNP distances from a single size of AuNPs ($d = 6.1 \pm 0.03$ nm) were used to calculate an average radial distance of 7.6 ± 0.6 nm (Figure 6-15D). This means that, using this simplified 2D geometry, the average distance from SWCNT surface to nanoparticle surface is 0.8 ± 0.3 nm.

We expand upon these 2D geometric analyses with 3D visualization of the ssDNA-AuNPs adsorbed on the SWCNT surface through *ab initio* modeling directly from scattering profiles using SASHEL (see SI Methods Section 6.5 and Extended Discussion Section 6.6). The final best-fit models demonstrate the complexity of these systems and validate our 2D interpretation (Figure 6-1E-F and Figure 6-16A-C). As expected, *ab initio* models for $(GT)_6$-AuNP-SWCNTs show increased AuNP packing density and greater consistency in inter-AuNP distances than $(GT)_{15}$-AuNP-SWCNTs. When rotated about the SWCNT axis, $(GT)_6$-AuNP-SWCNTs reveal a plane where there is little radial variance between AuNPs (Figure 6-1F), confirming our hypothesis that the small diameter of the SWCNT would reasonably support a 2D estimate of

geometries. Conversely, no such plane was found for $(GT)_{15}$-AuNP-SWCNT models and the AuNPs seem to rotate freely around the SWCNT axis (Figure 6-1E).

Our calculated separation distances of ssDNA polymers along the SWCNT axis are in agreement with previous literature.[244] Based on prior MD simulations, a single $(GT)_{15}$ polymer footprint on a (9,4) chirality SWCNT is expected to extend ~4 nm in length, with a ~2 nm helical pitch.[168] Another MD simulation-based study similarly estimates the pitch of $(GT)_{30}$ oligonucleotides on (11,0) SWCNTs to be 2-8 nm, depending on the DNA backbone orientation (with the 8 nm pitch orientation more energetically favorable, albeit on larger diameter SWCNTs).[245] This latter modeling study determined that a previously measured 18 nm pitch helix of poly-(GT) strands around SWCNTs via atomic force microscopy (AFM)[246] was structurally unstable and most likely introduced as an artifact during the air-drying step necessary for AFM sample preparation. Another AFM-based study suggests ~14 nm pitch for $(GT)_{15}$ on SWCNTs.[194] DNA pitch on SWCNTs has also been visualized by TEM, with estimates of 2.2 nm pitch for double-stranded salmon testes DNA along SWCNTs.[247] These previous pitch estimates can be converted to inter-strand spacing as measured in the current study via geometrical calculation, with an average SWCNT diameter of 1 nm and 0.7 nm length per ssDNA base,[194,231] assuming that inter- and intra-strand pitch distances are equivalent. Additionally, adjacent ssDNA strands are assumed to be close, but not intertwined, along the SWCNT axis.[231] Pitch estimates from the aforementioned previous literature ranging from 2-18 nm correspond to $(GT)_{15}$ inter-strand spacing of 15.7-20.9 nm and $(GT)_6$ inter-strand spacing of 6.3-8.4 nm. The directly measured inter-strand values in solution of 14.3 ± 1.1 nm and 8.6 ± 0.3 nm for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT, respectively, are reasonable in comparison to those computationally predicted or measured in previous studies.[244,248]

6.3.3 *Surface-adsorbed ssDNA structural changes as a function of ionic strength*

We applied this XSI approach to determine *in situ* ssDNA packing on the SWCNT surface as a function of solution ionic strength (Figure 6-2). Increasing solution ionic strength is expected to modify the surface-adsorbed ssDNA conformation, and thus AuNP scattering periodicity, by screening the negatively charged phosphate backbone of the ssDNA and enabling closer packing along the nanotube surface.[249,250] To test this hypothesis, ssDNA-AuNP-SWCNTs were synthesized, dialyzed against 0.1X PBS, and then diluted to various PBS concentrations to achieve different net salt concentrations while maintaining constant pH (see SI Methods Section 6.5). We tested samples in a range of 0.05X to 2X PBS represented as corresponding Debye lengths (λ_D) ranging from 3.37 to 0.53 nm, calculated as previously described.[251] This range was selected because ssDNA-SWCNTs are less stable in pure water and PBS concentrations above 2X resulted in aggregation.

As predicted, the longer, multi-pass helices of $(GT)_{15}$ on the nanotube surface compress at increased salt conditions (lower λ_D) from inter-AuNP distances of 14.7 to 12.3 nm (Figure 6-2A, C, and E). Conversely, the spacing of the shorter, single-pass rings of $(GT)_6$ did not change as significantly with ionic strength in either diagonal or axial inter-AuNP distances (Figure 6-2B, D, and E). An axial inter-AuNP shift from 17.5 to 16.8 nm was measured, corresponding to inter-ssDNA distances of 8.8 to 8.4 nm (for $\lambda_D = 3.37$-0.53 nm). Conversely, the calculated radial

distances show negligible change as a function of ionic concentration (for $\lambda_D = 3.37\text{-}0.53$ nm; Figure 6-15D).
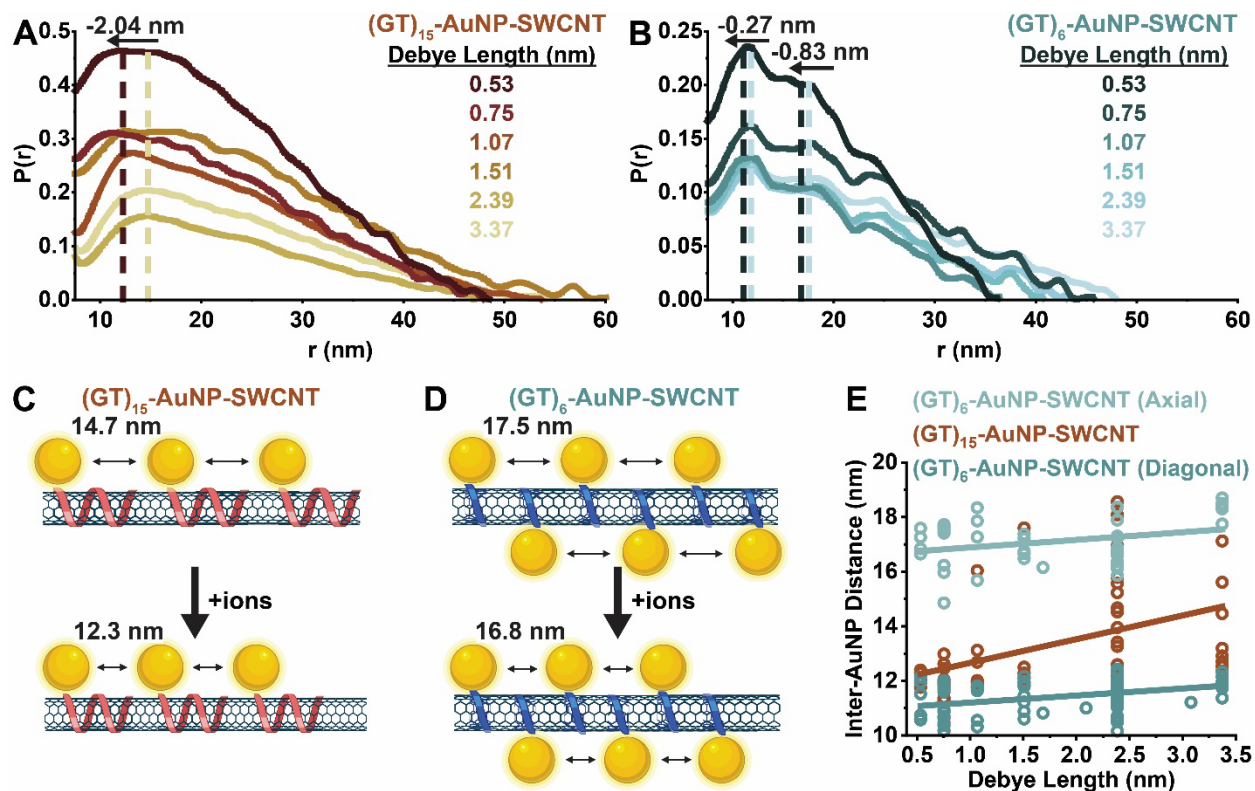
Figure 6-2. Surface adsorbed inter-ssDNA distance is modulated as a function of ionic strength for longer polymer lengths but remains relatively unchanged for shorter polymer lengths. Representative pairwise distribution functions, *P(r)*, for (A) (GT)$_{15}$-AuNP-SWCNTs (red-orange series) and (B) (GT)$_6$-AuNP-SWCNTs (blue series) in phosphate-buffered saline of varying net salt concentration, as represented by Debye lengths ($\lambda_D = 3.37\text{-}0.53$ nm). Dashed vertical lines are added to visualize peak shifts proceeding from light to dark dashed lines. *P(r)* functions are normalized to the primary intra-AuNP peak, then the x-axis minimum is set to focus on the inter-AuNP peak for clarity. (C-D) Schematic representations of changes in inter-AuNP distances at elevated ion concentrations for (C) (GT)$_{15}$-AuNP-SWCNTs and (D) (GT)$_6$-AuNP-SWCNTs. Schematics are not drawn to scale. Additional, representative *P(r)* functions and scattering curves are included in Figure 6-8 and Figure 6-9 for (GT)$_{15}$- and (GT)$_6$-AuNP-SWCNTs, respectively. (E) Summary of inter-AuNP distances as a function of Debye length ($\lambda_D = 3.37\text{-}0.53$ nm) for individual samples (dots) with corresponding linear regression (lines).

Given the smaller decrease in inter-AuNP distance observed for (GT)$_6$-AuNP-SWCNT at increased ionic strengths compared to that of (GT)$_{15}$-AuNP-SWCNT, we postulate that the high-salt condition affects the local intra-strand pitch to a greater extent than the neighboring inter-strand interactions. Moreover, increasing the concentration of ions in solution does not alter the

radial distances of $(GT)_6$-AuNPs across the SWCNT, as expected due to the short-range nature of the π-π interactions between the ssDNA and SWCNTs. Our salt-dependent ssDNA surface-packing results for $(GT)_{15}$ spacing on SWCNTs are in line with previous literature demonstrating this phenomenon with longer ssDNA on SWCNTs via indirect optical measurement and dried-state characterization.[249,250] At high salt concentrations, $(GT)_{30}$ was determined to adopt a compact conformation with higher SWCNT surface coverage,[249] putatively due to self-stacking of nucleobases from a related MD study.[252] In comparison, the ssDNA enters an elongated and stiffer conformation at low salt concentration, accompanied by ssDNA desorption from the SWCNT reducing the packing density.[249]

6.3.4   ssDNA-SWCNT nanosensor interactions with dopamine

We employed XSI to explore the ssDNA conformational changes of $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT complexes in the presence of the nanosensor target analyte, dopamine (DA) (Figure 6-3). Upon injection of 100 µM DA, a shift in the average inter-AuNP distances was observed in the $P(r)$ functions for both ssDNA lengths. Inter-$(GT)_{15}$ strand spacing increased by 2.11 ± 1.0 nm (Figure 6-3A, C, and Figure 6-17A), while the axial inter-$(GT)_6$ strand spacing increased by only 0.59 ± 0.27 nm (as calculated from the axial inter-AuNP peak shift of 1.17 ± 0.55 nm) and the diagonal inter-AuNP peak revealed an average shift of -0.93 ± 0.11 nm (Figure 6-3B-C and Figure 6-17B). Based on this observation, we calculated a corresponding radial inter-AuNP distance shift of -2.2 ± 0.41 nm for $(GT)_6$-AuNP-SWCNTs in the presence of DA, reducing the average SWCNT-to-AuNP-surface distance to -0.07 ± 0.04 nm. This dramatic decrease in radial distance demonstrates that DA causes the ssDNA to constrict around the SWCNT, drawing in the AuNP tags. Moreover, this slightly negative SWCNT-to-AuNP distance suggests that the AuNPs begin to overlap in the plane of the SWCNT and thus the ssDNA rings may be preferentially wrapping in opposite directions. Of note, this shortening of the radial inter-AuNP distances is not observed as a function of ionic strength (Figure 6-15D) and underscores the analyte-specific binding capabilities of this surface-constrained ssDNA sequence. *Ab initio* modeling of the AuNPs adsorbed on the SWCNT surface confirm the decrease in the average radial inter-AuNP distances for $(GT)_6$-AuNP-SWCNTs in the presence of DA (Figure 6-3D and Figure 6-18).

From the axial expansion of $(GT)_n$ down the length of the SWCNT and additional radial constriction of $(GT)_6$ onto the SWCNT, we postulate that DA both interacts with the phosphate groups of the ssDNA and inserts between ssDNA bases, depending on the initial conformation of the adsorbed polymers. Previous MD simulations of (8,8) chirality $(GT)_{15}$-SWCNTs (d = 1.1 nm) in the presence of DA suggest that the hydroxyl groups of DA (protonated at pH 7.4) interact with the exposed phosphate groups of the ssDNA backbone, drawing the ssDNA closer to the SWCNT surface.[221] Another MD study of (9,4) chirality $(GT)_{15}$- and $(GT)_6$-SWCNTs (d = 0.92 nm) reported that DA inserted between bases of the helically wrapped $(GT)_{15}$, but failed to insert between the bases of the ring-like $(GT)_6$, forming bridges between neighboring strands instead.[168] Both proposed mechanisms show that the DA interaction creates localized perturbations in the periodically ordered, ssDNA-induced electrostatic surface potentials of the SWCNT. These perturbations modulate exciton recombination lifetimes and lead to a large increase in nanosensor fluorescence. Our observation that the axial inter-strand distances of
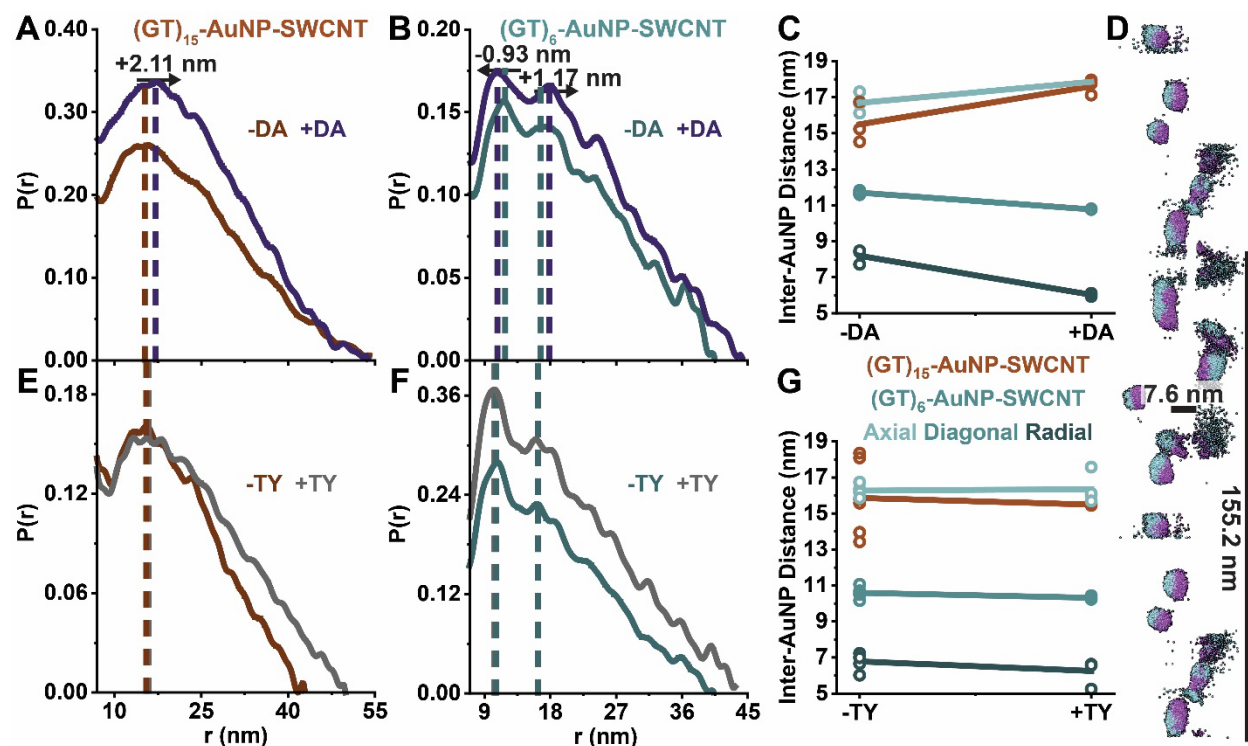
Figure 6-3. Inter-ssDNA distances shift in opposite directions for axial and radial spacing in the presence of dopamine and vary based on the ssDNA length and conformation. Inter-AuNP spacings shift in the presence of (A-D) dopamine (DA) but not with (E-G) *p*-tyramine (TY), a structural analog. Representative pairwise distribution functions, *P(r)*, with no analyte for (A and E) (GT)$_{15}$-AuNP-SWCNTs (red-orange series) and (B and F) (GT)$_6$-AuNP-SWCNTs (blue series) or in the presence of (A-B) DA (purple series) or (E-F) TY (grey series). Dashed vertical lines are added to visualize peak shifts. *P(r)* functions are normalized to the primary intra-AuNP peak, then the x-axis minimum is set to focus on the inter-AuNP peak for clarity. (C and G) Summary of inter-AuNP distances for replicates with and without (C) dopamine (DA) and (G) *p*-tyramine (TY). For (GT)$_6$-AuNP-SWCNT samples, axial, diagonal, and radial inter-AuNP distances are shown for individual samples (dots) with corresponding linear regression (lines). Corresponding *P(r)* functions for replicates are shown in Figure 6-17. (D) *Ab initio* modeling results for (GT)$_6$-AuNP-SWCNTs demonstrate the decrease in radial inter-AuNP distances as they move from blue to purple locations in the presence of DA. Fit and residuals are shown in Figure 6-18.

(GT)$_{15}$-AuNP-SWCNTs increase to a greater extent than (GT)$_6$-AuNP-SWCNTs complements the hypothesis that DA preferentially inserts between the bases for (GT)$_{15}$, increasing the pitch and hence footprint length along the SWCNT surface. Interestingly, as observed in (GT)$_6$-AuNP-SWCNTs, the presence of DA also constricts the ring-like structure around the SWCNT, suggesting that DA interaction is pulling the phosphate backbone of ssDNA towards the SWCNT surface. This interaction may also be the case for (GT)$_{15}$-SWCNTs, but no radial distances can be calculated due to the lack of orientational reference. As a control, XSI was

collected for ssDNA-AuNP-SWCNTs in the presence of the dopamine analogue, *p*-tyramine (TY), containing only one hydroxyl group. Negligible changes in the inter-ssDNA distances were observed upon injection of TY (Figure 6-3E-G and Figure 6-17C-D), consistent with the lack of fluorescence response and predictions from MD simulations.[166,168]

6.4   Conclusion

In this work, we demonstrate that XSI is a valuable technique for studying (GT)$_n$-SWCNTs (n = 6, 15) in solution, using small AuNP tags conjugated to the ssDNA to act as molecular rulers. XSI harnesses the tightly packed, electron-rich gold atoms in AuNPs to enable the study of nanomaterials at concentrations relevant to biological applications (0.1-5 mg/L). We find periodic ordering of ssDNA-AuNPs along the SWCNT axis, with the highest probability inter-AuNP distance of 14.3 ± 1.1 nm for (GT)$_{15}$-AuNPs and 11.4 ± 0.6 nm for (GT)$_6$-AuNPs. For (GT)$_6$-AuNP-SWCNTs, higher-order features observed after the most probable inter-AuNP distance motivated more detailed geometric calculations, giving rise to an extrapolated average inter-ssDNA ring distance of 8.6 ± 0.3 nm and an average distance from SWCNT-to-AuNP surface of 0.8 ± 0.3 nm.

Exploration of *in situ* ssDNA packing on the SWCNT surface as a function of solution ionic strength over the range of 0.05X to 2X PBS (λ_D = 3.37-0.53 nm) reveals an inter-ssDNA spacing decrease for (GT)$_{15}$-AuNP from 14.7 to 12.3 nm and (GT)$_6$-AuNP from 8.8 to 8.4 nm. These observations reflect the predicted electrostatic charge screening of the ssDNA phosphate backbone to permit closer packing. From these findings we postulate that the high-salt condition has a greater effect on the local intra-strand pitch rather than on the neighboring inter-strand interactions. The minimal change in radial AuNP spacing as a function of solution ionic strength suggests a lesser role of electrostatics in driving ssDNA-SWCNT adsorptive interactions, as expected for the likely π-π and/or other hydrophobic forces governing the polymer-surface adsorption mechanism.

XSI elucidates the conformational changes of (GT)$_{15}$ and (GT)$_6$ ssDNA adsorbed on SWCNTs in the presence of the nanosensor target analyte, DA, and provides insight into the mechanism responsible for the increased fluorescence response. Prior to analyte addition, radiative recombination of excitons is quenched by closely packed (GT)$_6$ rings that create periodic positive and negative surface potential pockets on the SWCNT.[168,253] Our observations demonstrate a perturbation of these predicted surface potentials in the presence of dopamine, as the axial distance is expanded between the strands and the ssDNA is pushed closer to the SWCNT surface, leading to an increase in radiative recombination pathways and/or decrease in nonradiative decay mechanisms. In the presence of DA, inter-ssDNA spacing increases by 2.11 ± 1.0 nm for (GT)$_{15}$-AuNPs and by only 0.59 ± 0.27 nm for (GT)$_6$-AuNPs. The greater shift in inter-ssDNA for (GT)$_{15}$-AuNP indicates that DA inserted between bases occur to a more significant degree for the helically wrapped (GT)$_{15}$, increasing the pitch and hence the footprint length on the SWCNT surface as previously predicted.[168] Interestingly, for (GT)$_6$-AuNP-SWCNTs the radial inter-AuNP distances show a dramatic decrease of 2.2 ± 0.41 nm in the presence of DA, which is not detected in response to changes in salt concentrations. This change is confirmed by *ab initio* modeling and demonstrates a constricting of the (GT)$_6$ ring likely caused by the hydroxyl groups of DA interacting with the exposed phosphate groups of the

ssDNA, pulling the ssDNA closer to the SWCNT surface. This constriction leads to a calculated SWCNT-to-AuNP surface distance of -0.07 ± 0.04 nm, implying that AuNPs overlap in the plane on one side of the SWCNT. This potential geometry suggests that the ssDNA rings have coordinated directionality, wrapping in opposite directions from each other along the SWCNT surface. These findings attest to the complexity of these nanobiotechnologies, suggesting that the mechanisms behind their molecular recognition of DA is conformationally driven, and hence sequence-specific, leading to a strong argument towards the need for rational design to properly tailor their optical properties.

As demonstrated, XSI provides a powerful tool to complement previously employed techniques to characterize DNA-based nanotechnologies under biologically relevant solution-phase conditions. Through this technique, we gain understanding of the discrete nanoscale architectures of these materials and their mechanisms of interaction with the local environment in solution. Such high-throughput measurements are important to understand how polymer-nanoparticle complexes function for a broad range of nanobiotechnology applications.

6.5   Materials and Methods

6.5.1   Synthesis of Citrate-Capped Gold Nanoparticles (AuNPs)

Citrate-capped AuNPs of diameters 5.9-7.2 nm were prepared using a method modified from that which was previously described.[134] Briefly, a 2 L solution of 0.25 mM $HAuCl_4$ and 0.25 mM tri-sodium citrate was prepared in a conical flask using ddH_2O cooled to 4°C. Next, 10 mL of 0.6 M $NaBH_4$ at 4°C was added rapidly to the solution while stirring. The solution turned dark red immediately after adding $NaBH_4$, indicating particle formation. The solution was allowed to warm to room temperature (RT) and stirred overnight to no reactivity from excess $NaBH_4$.

6.5.2   Citrate-BSPP Exchange for Gold Nanoparticles (BSPP-AuNPs)

Bis-(p-sulfonatophenyl) phenylphosphine (BSPP; Millipore Sigma, Burlington, MA) was added to citrate-stabilized colloidal AuNPs (~5.0 x 10^{13} particles/mL) to a final concentration of 0.5 g/L and stirred at RT for a minimum of 6 hours. Approximately 1 mL of saturated NaCl solution was added per 10 mL BSPP-exchanged colloidal gold, until the solution changed from transparent red to a darker, cloudy purple, indicating the reversible precipitation of the AuNPs.[135] The mixture was centrifuged (Beckman Coulter, Avanti J25, JA-18 rotor, Indianapolis, IN) at 12,000 rcf for 10 min and supernatant decanted to waste. BSPP-AuNPs were washed twice with 0.5 M NaCl solution (repeating the centrifugation step above) and resuspended in 15 mM phosphate buffer, 1 mM TCEP, pH 7 for storage. Suspensions were stored at 4°C until use. Note that freezing caused sample precipitation.

6.5.3   Conjugation of Single-Stranded DNA to BSPP-AuNPs (ssDNA-AuNPs)

If stored longer than 2 weeks, fresh TCEP was added to reduce the solution of colloidal BSPP-AuNPs prior to ssDNA conjugation by adding saturated NaCl solution to BSPP-AuNPs until the solution turned dark/cloudy, centrifuging at 12,000 rcf for 10 min, and resuspending in fresh 15

mM phosphate buffer, 1 mM TCEP, pH 7. The final BSPP-AuNP concentration was determined by measuring the absorbance at 520 nm (NanoDrop 2000, Thermo Scientific) and converting to concentration with the empirical extinction coefficient,[145] $\varepsilon_{520nm} = 9.69 \times 10^6$ L mol^{-1} cm^{-1}. The concentration of desired oligonucleotides was calculated by measuring the absorbance at 260 nm using the sequence-dependent extinction coefficient. BSPP-AuNPs were attached to ssDNA via trithiolated linkers (Letsinger's type) on the 5'-end of ssDNA oligomers with SDS-PAGE purification (Fidelity Systems, Gaithersburg, MD) by mixing at a 1:1 molar ratio and incubating at RT overnight. Conjugated ssDNA-AuNPs were then coated with short, neutral Methoxy poly(ethylene glycol) thiol (mPEG-SH; MW ~350 g/mol; Biochempeg Scientific Inc., Watertown, MA), by adding mPEG-SH to the ssDNA-AuNP suspension at a final molar ratio of 3000:1 mPEG-SH to AuNPs.

6.5.4   Anion Exchange Chromatography Purification of ssDNA-AuNPs

Mono-conjugated ssDNA-AuNPs were isolated as previously described[12,136] using a Dionex DNA-Pac PA100 anion exchange column on either a GE AKTA Explorer or a GE Atka Pure fast protein liquid chromatography (FPLC; General Electric HealthCare Technologies, Chicago, IL) with an NaCl gradient from 0.01 to 1 M over a period of 55 min at a flow rate of 1.1 mL/min. Sample elution was monitored by measuring UV-Vis absorption at 260 (ssDNA) and 520 nm (AuNPs), and the mono-conjugated ssDNA-AuNP fraction was collected for downstream use (Figure 6-5A). The same AuNP conjugation and purification method was implemented for both (GT)$_{15}$ and (GT)$_6$ oligomers, with the longer demonstrating increased retention times via anion exchange chromatography, as expected (Figure 6-5B). The final ssDNA-AuNP concentration was determined again by measuring absorbance.

6.5.5   Data Collection and Processing by High-Throughput X-ray Scattering Interferometry (HT-XSI)

HT-XSI data was collected at the SIBYLS beamline (bl12.3.1) at the Advanced Light Source of Lawrence Berkeley National Laboratory, Berkeley, California.[2] X-ray wavelength was set at $\lambda = 0.12398$ nm and the sample-to-detector distance was 2.07 m, resulting in a scattering vector (q) range of 0.1 - 4.6 nm^{-1}, which corresponds to real-space distances of 62.8 - 1.4 nm. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, with scattering angle 2θ. Data was collected using a Dectris PILATUS3X 2M detector and processed as described previously.[10]

Immediately prior to data collection, 15 µL of each sample was added to 15 µL of buffer in a 96-well plate for final corresponding concentrations of 42.5 - 450 nM ssDNA-AuNPs and 0.17 - 1.76 mg/L SWCNTs. Each sample was then transferred to the XSI sampling position via a Tecan Evo liquid handling robot (Tecan Trading AG, Switzerland) with modified pipetting needles acting as sample cells as described previously.[43] Samples were exposed to X-ray synchrotron radiation for 5 s at a 0.1 s frame rate for a total of 50 images. Each collected image was circularly integrated and normalized for beam intensity to generate a one-dimensional scattering profile. Buffer subtraction was performed for the one-dimensional scattering profile of each sample using each of two bracketing buffer wells to ensure the subtraction process was not subject to instrument variations. Scattering profiles over the 5 s exposure were sequentially

averaged together to eliminate any potential radiation damage effects. All data processing was done using our beamline specific data processing pipeline by the SIBYLS SAXS Process (SSP) GUI (see Sections 2.6 and 3.4).[12,42] Pairwise distribution functions, *P(r)*, were generated in batch using the automated GNOM[44] feature of the SSP GUI.

6.5.6 Characterization by Small Angle X-ray Scattering (SAXS)

Average SWCNT bundling of ssDNA-SWCNTs (no AuNPs) determined by calculating the cross-sectional radius of gyration (R_{cs}) for rod-like scatterers from SAXS curves using ATSAS 3.0.[92] Absolute-scale intensity scattering measurements and calculations were completed as previously described.[254] All AuNPs used in this study are near-spherical, with average diameters ranging from 5.9 to 7.2 nm determined from pairwise distribution functions, *P(r)*, obtained from SAXS curves for each batch of synthesized ssDNA-AuNPs (Figure 6-6A, Table 6-1, and Table 6-2). SAXS profiles for all ssDNA-AuNPs were further modeled as triaxial ellipsoidal fittings[137] using SasView software to obtain more accurate nanoparticle dimensions (Figure 6-6B, Table 6-1, Table 6-2, and detailed in SI Extended Discussion Section 6.6). The average ratio of largest (major equatorial radius, r_A) to smallest (polar radius, r_C) particle dimension is observed at 3.3:2. Including a polydispersity parameter for r_A was essential to optimize the fit (Figure 6-6C). The average polydispersity index (PDI) of r_A is 0.18, indicating reasonable monodispersity, where PDI < 0.1 is considered ideal.[138]

6.5.7 Electron Density Calculations

For X-ray scattering experiments, the electron density of a material is of critical importance, as the total scattering intensity is proportional to the square of the electron density. The triaxial ellipsoidal AuNPs used in this study have a calculated electron density of 3519.08 e⁻/nm³, increasing the electron density of our otherwise low-scattering materials with calculated electron densities of 914.3 and 1818 e⁻/nm³ for SWCNTs and ssDNA, respectively. Detailed electron density calculations are found in SI Extended Discussion Section 6.6.

6.5.8 Characterization by Dynamic Light Scattering (DLS)

DLS measurements were taken with the Zetasizer Nano ZS (Malvern Analytical) with a material refractive index of 0.200 and absorption of 3.320 for colloidal gold.[151 and Aluminum. I. The Absorption Coefficient k] All samples were diluted in 0.1X PBS to an AuNP concentration of 0.20-0.25 µM and loaded in disposable cuvettes (Malvern ZEN0040) for size measurement. AuNPs were coated with mPEG-SH to prevent aggregation and aid in the purification by anion exchange chromatography.[136] After mPEG-SH coating, the hydrodynamic radii of the AuNPs measured by DLS increase from the AuNP core radii measured with SAXS by an average of 1.46 ± 0.34 nm across all samples (Figure 6-6D and Table 6-3). This increase in hydrodynamic diameter implies successful mPEG-SH functionalization of the exposed AuNP surfaces.[146] The average length of the mPEG-SH (MW ~350 g/mol) is ~1.7 nm when fully extended, based on the length of the repeating polyethylene oxide unit of 0.278 nm in water (with n=6 units total).[147] However, even if all available mPEG-SH functionalized the surface, the theoretical distance between grafting

sites (~5.2 nm) would far exceed the calculated Flory radius of 0.815 nm, leading to only a partially extended mushroom conformation. This result is in line with the measured change in hydrodynamic radius imparted by mPEG-SH being slightly less than the expected change based on polymer length.[148-150]

6.5.9   Suspension of Single-Walled Carbon Nanotubes (SWCNTs) with ssDNA-AuNPs

ssDNA-AuNP concentration was determined by measuring the absorbance at 520 nm (NanoDrop One, Thermo Scientific) with a 10X-diluted aliquot and calculating the concentration as before. Single-walled carbon nanotube (SWCNTs) were suspended with ssDNA-AuNPs as follows: mixed-chirality raw SWCNTs (small diameter HiPco™ SWCNTs, raw, NanoIntegris, Boisbriand, Quebec, Canada) were first prepared as an aqueous slurry of 2 mg/mL in Milli-Q water. ssDNA-AuNP-SWCNTs were then formulated to maintain a final ratio of 250 nmol ssDNA-AuNP per 1 mg SWCNT, at a total volume of 2-4 mL (such that half of the solution could serve as non-sonication controls). The exact formulation recipe depended on the yield of ssDNA-AuNPs obtained after anion exchange purification. For every 2 mL of ssDNA-AuNPs at 200-800 nM (0.4-1.6 nmol), 0.8-3.2 µL of SWCNT slurry was added (1.6-6.4 µg) in 0.1X phosphate-buffered saline (PBS; note 1X PBS is 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) in a 5 mL tube. The ssDNA-AuNP/SWCNT mixture was bath sonicated for 10 min (Branson Ultrasonic 1800) then probe-tip sonicated for 10 min in an ice bath (3 mm probe tip at 50% amplitude, 5-6 W, Cole-Parmer Ultrasonic Processor). ssDNA-AuNP-SWCNT suspension was equilibrated for 30 min at RT then dialyzed against 2 L of 0.1X PBS overnight (200 µL volume in Pur-A-Lyzer Mini Dialysis Kit with 6-8 kDa MWCO, Millipore Sigma). Note that free ssDNA-AuNPs are not expected to pass through this filter size that contains pores of only a few nanometers and this step was included as a buffer exchange to remove any impurities still present from the AuNP synthesis. Suspensions were stored at 4°C until use. Control experiments without SWCNTs were prepared from the same batch of ssDNA-AuNPs, with all steps the same but in the absence of SWCNTs.

Certain parameters were slightly modified in comparison to usual ssDNA-SWCNT suspension protocols to account for the AuNP tag on the ssDNA. First, the material amounts were reduced approximately two orders of magnitude to account for the limited availability of AuNPs, although kept in a similar ratio to previous suspension protocols (250 nmol ssDNA:1 mg SWCNT).[166,238] Second, a post-sonication pelleting step, which one would typically perform in order to remove unsuspended SWCNTs or amorphous carbon and catalyst left over from SWCNT synthesis, was omitted because the presence of the AuNPs causes full sample pelleting due to the additional mass. We expect that reducing the overall material load by two orders of magnitude also reduced the concentration of SWCNT-derived impurities in the final suspension, thus eliminating the need for a centrifugation clean-up step. Finally, SWCNT suspensions were not spin-filtered due to embedding of the AuNPs into the filter membrane and full sample loss. Unless specified, all reagents, including Carboxylated SWCNTs used for controls, were purchased from Millipore Sigma (Burlington, MA). Additional suspension notes and controls are detailed in SI Extended Discussion Section 6.6.

6.5.10 Characterization by Absorbance and Fluorescence

Absorbance was measured with a UV-VIS-nIR spectrophotometer (UV-3600 Plus, Shimadzu Corporation, Kyoto, Kyoto, Japan) using 50 μL sample volume in black-sided quartz cuvette (Thorlabs, Inc., Newton, NJ). Near-infrared SWCNT fluorescence was measured using an inverted Zeiss microscope (Axio Observer.D1, 10x objective, Carl-Zeiss-Stiftung, Oberkochen, Baden-Württemberg, Germany) with a Princeton Instruments spectrometer (SCT 320) and liquid nitrogen-cooled Princeton Instruments InGaAs detector (PyLoN-IR; Teledyne Technologies, Thousand Oaks, CA). A triggered 721 nm laser (OptoEngine LLC, Midvale, UT) was used as the excitation source and fluorescence emission was collected from 800 – 1400 nm. 30 μL volume of each sample was prepared in polypropylene 384 well-plates (Greiner Bio-One microplate).

6.5.11 Characterization by Transmission Electron Microscopy (TEM)

Images of $(GT)_6$-AuNP-SWCNT and $(GT)_{15}$-AuNP-SWCNT complexes were captured using a Tecnai 12 TEM (FEI, Hillsboro, OR) operating at an accelerating voltage of 120 kV and data was recorded using a Gatan Rio16 CMOS camera with GWS software (Gatan Inc., Pleasanton, CA). Samples were prepared by depositing 5 μL of sample onto 400 mesh carbon/formvar-coated copper grids (EMS Electron Microscopy Science) that were surface treated by glow discharge to render the support hydrophilic. The samples were wicked away after 2 minutes. No negative staining or washing steps were included. Approximately 20 images were taken at four different regions on the grid for each sample to ensure reported images were representative. Additional image analysis was performed using Fiji (Imagej2).[255] To determine the size distribution and average AuNP size, over 1000 AuNPs were analyzed. To calculate the AuNP packing on the SWCNTs, a minimum of 5 SWCNTs were analyzed per sample giving a total SWCNT length of 2980 ± 62 nm.

6.5.12 Ab initio Modeling of XSI Data

SASHEL, adapting the methodology originally described by Burian and Amenitsch.[256] SASHEL was first developed as an algorithm to reconstruct helical and rod-like systems by randomly moving dummy atoms comprising a single building block unit and projecting them outward using symmetrical boundary conditions. Theoretical scattering profiles from these projected models are iteratively fit to experimental data to converge on the best-fit model. In a typical SASHEL analysis, it is recommended to reduce the number of total datapoints uniformly across the original 1D curve (i.e., removing every other datapoint) to lower calculation times. However, this leads to a higher density of datapoints in the high q-range due to the binning of the circular integration of the detector images and reduces the fitting quality in the low q-range. In this study, datapoints were sequentially removed at higher density as the q-value increases to retain high fit quality in the low q-range, corresponding to the longer order distances of interest. The data was fit over a truncated q-range of 0.1-3 nm^{-1} to further focus on the lower q-range.

Initial conditions for the model were refined over many iterations, with the most robust outcome derived from a carefully curated starting model for both $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT samples (Figure 6-19). First, the more clearly defined $(GT)_6$-AuNP-SWCNT scattering curve

was used to produce a crude initial estimate of dummy-atom positions (Figure 6-19A). This model started with a core-shell cylinder model with 15 nm outer diameter and 5 nm inner diameter, as determined by the average diameter of the SWCNT (~1 nm) together with a fully extended trithiolated linker on either side (~2 nm each). The stack building block height (H_{BB}) was set equal to the axial inter-AuNP distance (17.2 nm) of $(GT)_6$-AuNP-SWCNT, as determined from $P(r)$ functions (Figure 6-1B). The number of stacks (N_S) was set to 15 to ensure an overall length far greater than the observed maximum length dimension, d_{max}, of the $(GT)_6$-AuNP-SWCNT samples (~55 nm). The starting temperature (T_0) was set to 0.6 to allow for broad movement of the 2000 initial dummy atoms per H_{BB}. T_0 in this case is the value in which the system starts to cool down as defined previously[256] and is not representative of temperature on an absolute scale. Dummy-atom diameters were set to 0.288 nm to simulate the atomic diameter of a gold atom. From this initial model, clusters of dummy atoms formed and were taken to be naturally representative of AuNPs (regions of high electron density). The most clearly defined cluster was extracted, duplicated to create a pair of AuNPs, and the number of dummy atoms expanded to 1000 for each AuNP (Figure 6-19A).

Each AuNP in the initial pair is maneuvered into estimated initial geometries as determined experimentally to represent the inter-AuNP block heights (H_{GB}) unique for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs (14.3 and 17.2 nm, respectively). Each refined AuNP pair (H_{GB} unit) was then replicated axially using symmetrical boundary conditions an integer (k) number of times to produce a total stack height, $H_{BB} = k*H_{GB}$, as depicted in Figure 6-19B and C for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs, respectively. After exploring a large parameter space, (outlined in detail in SI Extended Discussion Section 6.6, Table 6-4, Table 4-1, Figure 6-20, Figure 6-21, Figure 6-22, and Figure 6-23) it was discovered that when so few dummy atoms were used per AuNP, the models would fit better by expanding outward to compensate for a lack of representative electron density. Thus, initially modeling $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs as two neighboring ssDNA-AuNP-SWCNTs saved a lot of computational time (fewer dummy atoms to move per iteration) and allowed for a larger parameter space to be explored. These initial models were run at a starting temperature of 0.2 at each of the respective H_{BB} values, with N_S = 32 / number of AuNPs (N_{NP}) rounded to the nearest integer value, where N_{NP} is the total number of AuNPs per stack. All models were run for 200 iterations, as the goodness of fit (χ^2-value) tends to reach a minimum plateau.

To model the system as a single SWCNT, the best fit double SWCNT model for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT was selected (Figure 6-24) and the single SWCNT region showing the clearest AuNP spacing is extracted. The regions of clear electron density are replaced with denser 3000 dummy-atom clusters (representing AuNPs; Figure 6-19B-C) and an initial H_{BB} was used which resulted in the best fit from each respective double SWCNT model (H_{BB} = 114.5 and 155.2 nm for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT respectively; Table 4-1 and Figure 6-23). The N_S were 3 and 2 for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT models, respectively, to obtain final models of similar total length (342.6 and 310.4 nm, respectively). Over many iterations, as regions in the model appeared which suggested areas of excessive or missing electron density, AuNPs (dummy-atom clusters) were manually removed or added accordingly until achieving a reasonably fitting single SWCNT model (χ^2-value < 2; Figure 6-16A-B). Higher starting temperatures (T_0 > 0.4) were generally used for single SWCNT to speed up the movement of initial models containing at least three-fold more dummy atoms than earlier models.

6.6   Extended Discussion

6.6.1   Electron Density Calculations

SWCNTs

The surface structure of a SWCNT is that of a sheet of graphene.[257] Assuming the C-C bond length $d_{c\text{-}c} = 0.1421$ nm is the same for graphene as for a curved SWCNT, the surface area of one carbon hexagon (a_h) is:

$$a_h = \frac{3\sqrt{3}}{2} d_{c-c}^2 = 0.0525 \; nm^2$$

6-1

The calculated surface area of one benzene ring corresponds to two carbon atoms and thus the electron surface density of the hexagon (σ) is:

$$\sigma = \frac{\text{number electrons } (e^-)}{a_h} = 228.57 \; e^-/nm^2$$

6-2

The electron density (ρ) of a SWCNT is calculated as follows with surface area (a_s) = πdL and volume (V) = $\pi (d/2)^2 L$ for a SWCNT of average diameter (d) = 1 nm (according to technical data sheet from NanoIntegris) and arbitrary length (L):

$$\rho = \frac{a_s \sigma}{V} = \frac{4\sigma}{d} = 914.3 \; e^-/nm^3$$

6-3

AuNPs

The number of gold atoms is based on calculations from ellipsoidal AuNPs, as previously described.[258] The equation for calculating the number of atoms in a triaxial ellipsoidal AuNP (N_{Au}) is as follows:

$$N_{Au} = \frac{4\pi}{3V_{Au}} (r_A - d_{Au})(r_B - d_{Au})(r_C - d_{Au})$$

6-4

Where $V_{Au} = 0.017$ nm³ is the volume of the gold atom and $d_{Au} = 0.288$ nm is the diameter of the gold atom at room temperature. r_A, r_B, and r_C define the major equatorial, minor equatorial, and polar radii of the triaxial ellipsoidal fit of the AuNPs[137] as noted in Figure 6-6, Table 6-1, and Table 6-2. The volume of an AuNP (V_{NP}) is calculated and follows:

$$V_{NP} = \frac{4\pi}{3} r_A r_B r_C$$

6-5

The electron density for AuNPs (ρ_{NP}) is then obtained as follows, with the number of electrons in a gold atom (N_e):

$$\rho_{NP} = \frac{N_e}{V_{Au}} \frac{(r_A - d_{Au})(r_B - d_{Au})(r_C - d_{Au})}{r_A r_B r_C}$$

6-6

Finally, with the electron density of a single gold atom $\rho_{Au} = N_e/V_{Au}$, Equation 6-6 simplifies to the following, which can be used to define the electron density for any ellipsoidal or spherical AuNP assuming a face-centered cubic (FCC) unit cell:

$$\rho_{NP} = \rho_{Au} \prod_{i=A,B,C} (1 - \frac{d_{Au}}{r_i}) = 3519.08 \, e^-/nm^3$$

6-7

(GT)$_n$ ssDNA

The volume for repeating units of (GT)$_n$ was estimated using the volume modeling feature in UCSF Chimera[259] and the number of electrons in the system was calculated directly by elemental composition.

6.6.2   Triaxial Ellipsoidal Fitting

Triaxial ellipsoidal fittings of AuNP scattering curves were performed using SasView by incorporating the classical physical properties of the light scattering of ellipsoidal particles in solution.[137] Here, the scattering of randomly ordered particles in solution is expressed as follows:

$$I(q) = scale(\Delta\rho)^2 \frac{V_{NP}}{4\pi} \int_{\Omega} \phi^2(qr) \, d\Omega + background$$

6-8

Where the integral sums all possible rotational orientations in solution, V_{NP} is defined in Equation 6-5, and $\Delta\rho$ is the scattering length density difference between the sample and buffer. For the triaxial ellipsoidal fittings in this work, the buffer was 0.1X PBS, with a scattering length density close to ρ_{water} at room temperature (9.44×10^{-6} Å^{-2}) used as the starting value of the fitting. Other fit parameters were left unconstrained and the DREAM algorithm[153] was used to fit the data. The fitting was further optimized by allowing for the polydispersity of the major equatorial radius (r_A) to be modeled as a Gaussian distribution, where the polydispersity index, PDI = standard deviation / mean (Figure 6-6C). Fittings are shown in Figure 6-6B and data are tabulated in Table 6-1 and Table 6-2.

6.6.3   ssDNA-AuNP-SWCNT Synthesis

Two alternative methods were attempted to suspend SWCNTs with ssDNA-AuNPs:

(1) **Biotin-streptavidin binding method**: Based on previous work,[249] SWCNTs (0.2 mg, NanoIntegris) were suspended with 3'-biotinylated $(GT)_6$ and $(GT)_{15}$ oligomers (1 mg, Integrated DNA Technologies) in 1 mL total volume of 0.1X PBS via the same probe-tip sonication protocol as described earlier. Sonication was followed by centrifugation at 16,100 rcf for 30 min and separation of the supernatant containing the ssDNA-SWCNT product. The ssDNA-SWCNT solution was concentrated by centrifugal filtration (100 kDa MWCO, Millipore Sigma) with one Milli-Q water wash step. The ssDNA-SWCNT suspension was then mixed at various ratios with streptavidin-functionalized AuNPs (5 nm diameter, Cytodiagnostics Inc.). Although the mixtures were still near-infrared fluorescent (indicating biotin-ssDNA suspension of SWCNTs), no complexes of ssDNA-AuNP-SWCNTs were observed via SAXS measurement. This method was also not pursued further due to the multivalency of streptavidin (each streptavidin molecule has four attachment points for biotin), resulting in the potential for SWCNT bridging and subsequent artifacts in analysis.

(2) **Corona exchange method**: Based on our previous work,[178] SWCNTs (0.2 mg, NanoIntegris) were suspended with untagged $(GT)_{15}$ oligomers (1 mg, Integrated DNA Technologies) in 1 mL total volume of 0.1X PBS via the same probe-tip sonication protocol as described earlier. Sonication was followed by centrifugation at 16,100 rcf for 30 min and separation of the supernatant containing the ssDNA-SWCNT product. The ssDNA-SWCNT solution was concentrated, and free ssDNA was removed by centrifugal filtration (100 kDa MWCO, Millipore Sigma) with five Milli-Q water wash steps. Exchange of the initial corona (untagged ssDNA) for the final corona (AuNP-tagged ssDNA) was attempted either by passive exchange or by dialysis. For the former case, $(GT)_{15}$-AuNPs were mixed with $(GT)_{15}$-SWCNTs in molar ratios ranging from 30-30,000 AuNP-tagged ssDNA to SWCNTs (10 total ratios tested, all in 1X PBS). This range of molar ratios encompasses and far exceeds the expected value of 140 $(GT)_{15}$ strands per SWCNT,[178] in the aim of providing a driving force for AuNP-tagged ssDNA adsorption onto SWCNTs. For the latter case of dialysis-based exchange, the $(GT)_{15}$-SWCNT suspension (60 mg/L) was mixed with free $(GT)_{15}$-AuNPs (74 nM of 5 nm diameter AuNPs or 132 nM of 7 nm diameter AuNPs) and overnight dialysis (~16 h against 2 L of Milli-Q water) was used to remove untagged, desorbed ssDNA (0.5 mL volume in Slide-A-Lyzer™ Dialysis Cassettes with 20K MWCO, Thermo Fisher). For both corona exchange attempts, no complexes of ssDNA-AuNP-SWCNTs were observed via SAXS measurement. This is most likely due to the lower exchange rate for the bulkier AuNP-tagged ssDNA onto the SWCNT surface in place of the smaller untagged ssDNA already on the surface.

Purification of resulting ssDNA-AuNP-SWCNT complexes was also attempted. The centrifugation step normally included after ssDNA-SWCNT suspension (16,100 rcf for 30-90 min to remove any SWCNTs not suspended by ssDNA in the pellet) caused complete sample sedimentation due to the large AuNP mass. Centrifugal filtering was also attempted (3 kDa MWCO, Millipore Sigma). However, the combination of AuNPs and SWCNTs led to complete and irreversible sample embedding within the membrane filter.

Finally, we explored the possibility of iron nanoparticle contamination in SWCNT samples accounting for the X-ray scattering results. Namely, the HiPco SWCNT synthesis process involves the *in situ* thermal decomposition of iron pentacarbonyl to iron catalyst nanoclusters upon which the SWCNTs then grow.[260] Raw HiPco SWCNTs are reported to contain <35 wt% residual iron catalyst (NanoIntegris characterization data). We have previously measured approximately 8.5 wt% iron present in our raw HiPco SWCNT samples once suspended with ssDNA, via X-ray photoelectron spectroscopy analysis (unpublished). To test the potential influence of iron nanoclusters, we further tested super-purified HiPco SWCNTs (NanoIntegris; reported to contain <5 wt% residual iron catalyst and presumably also less once ssDNA-suspended) similarly suspended in ssDNA. Baseline X-ray scattering at the relevant SWCNT concentration with these super-purified samples was negligible in comparison to that in the presence of AuNPs, supporting the insignificant contribution of residual iron catalyst in the observed scattering profiles.

6.6.4   *Ab initio modeling of ssDNA-AuNP-SWCNT complexes*

We adapted an *ab initio* modeling technique to produce 3D models complementing our 2D description of the AuNP geometries on the SWCNT surface. A large parameter space of initialization geometries was explored to converge on best-fit *ab initio* models for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT complexes. In structural biology, SAXS profiles from macromolecules are traditionally modeled via *ab initio* shape reconstruction using software such as DAMMIN[261], GASBOR[262], or more recently, DENSS[263]. However, these 3D modeling techniques all fail to capture samples with very high aspect ratios, such as SWCNTs. 3D modeling of the AuNPs adsorbed on the SWCNT surface was instead accomplished using the *ab initio* modeling capabilities of SASHEL, expanding upon the methodology described by its developers Burian and Amenitsch.[256] SASHEL was adapted to move clusters of 1000 dummy atoms, where each cluster represents a single AuNP, starting from curated initial geometries (see SI Methods Section 6.5 and Figure 6-19. Schematic demonstration of adapted SASHEL *ab initio* modeling strategy with labeling of important definitions for stack building block heights (*HBB*), inter-AuNP block heights (*HGB*), number of stacks (*NS*), and number of AuNPs (*NNP*). The stepwise methodology is as follows: (A) broad movement of 2000 initial dummy atoms from initial guess core-shell model (green) resulting in a densely packed strand of dummy-atom clusters, each representative of an AuNP (pink). (i) One clearly defined cluster is selected and duplicated to give an AuNP pair, (ii) then the pair of AuNPs is expanded to 1000 dummy atoms per AuNP (purple). Each AuNP in the pair is maneuvered into desired initial geometries as determined experimentally, representing one *HGB* unit, unique for (B) $(GT)_{15\text{-AuNP-SWCNTs}}$ and (C) $(GT)_{6\text{-}}$ $_{\text{AuNP-SWCNTs.}}$ (iii) The *HGB* is replicated in one direction using symmetrical boundary conditions an integer (k) number of times to produce a total stack height, $HBB = k*HGB$, using parameters in Table 4-1. (iv) The starting model is then used to produce ab initio models representing two neighboring ssDNA-AuNP-SWCNTs. (v) The single ssDNA-AuNP-SWCNT showing the clearest AuNP spacing is selected and the regions of electron density are replaced with denser 3000 dummy-atom clusters (AuNPs).A). The best-fit models obtained using this technique demonstrate the dynamic nature of the system, providing a larger graphical view of the system as it exists in solution and complementing our statistical analysis of inter-AuNP distances.

It was initially determined that both $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs were best captured by parallel lines of AuNPs (SWCNTs) starting at a distance of 15 nm apart (Table 6-4 and Figure 6-20). Based on our statistical analysis, the average radial inter-AuNP distance for $(GT)_6$-AuNP-SWCNTs is 7.6 ± 0.6 nm. Thus, *ab initio* modeling results demonstrate distances over two-fold higher, indicating that two neighboring ssDNA-AuNP-SWCNTs are required to properly describe the overall volume of the electron density of the samples. The starting stack-building block heights (H_{BB}) and hence the initial number of AuNPs (N_{NP}) ranging from 2 to 18 AuNPs were attempted in order to expand the volume of the modeling space and modulate complexity within the modeling (Table 4-1, Figure 6-21, and Figure 6-22). The H_{BB} range attempted represents integer multiples of the inter-AuNP block heights (H_{GB}) for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT (14.3 and 17.2 nm, respectively), as determined by the average axial inter-AuNP distances from $P(r)$ functions. Both $(GT)_{15}$ and $(GT)_6$-AuNP-SWCNTs demonstrated local minima in χ^2-values when plotted as a function of H_{BB}, with average minima at integer multiples of 54.5 ± 2.8 nm (Figure 6-23). This H_{BB} value could represent the upper end of the observed maximum length dimension, d_{max}, range of the ssDNA-AuNP-SWCNT samples (~45-55 nm) as determined from $P(r)$ functions (Figure 6-1A-B). $(GT)_{15}$ and $(GT)_6$-AuNP-SWCNTs were best fit using starting models Containing 16 AuNPs with H_{BB} of 114.5 nm and 18 AuNPs with H_{BB} of 155.2 nm, respectively (Table 4-1 and Figure 6-24A-B). Initially, these AuNPs are evenly arranged according to experimental geometries (see SI Methods Section 6.5 and Figure 6-19), leading to calculated AuNP packing of 0.14 and 0.12 AuNP per nm for $(GT)_{15}$ and $(GT)_6$-AuNP-SWCNT, respectively. While the calculated AuNP packing of $(GT)_{15}$-AuNP-SWCNT complemented those calculated from TEM analysis (0.139 AuNPs per nm length of SWCNT), those calculated for $(GT)_6$-AuNP-SWCNT were inconsistent with the corresponding 0.185 AuNPs per nm length of SWCNT from TEM analysis. Interestingly, the results of the *ab initio* modeling for $(GT)_{15}$-AuNP-SWCNTs maintain a consistent number of AuNPs, while for $(GT)_6$-AuNP-SWCNTs some dummy atoms split off from the original AuNPs and start to form new AuNPs, suggesting that there are regions of missing electron density that necessitate additional AuNPs. For purposes of comparison, the single ssDNA-AuNP-SWCNT showing the most clearly defined AuNPs was isolated from each model and compared to theoretical models produced using Solidworks (Dassault Systèmes) directly from geometries obtained from statistical analysis of the $P(r)$ functions (Figure 6-24C). As predicted, the isolated $(GT)_{15}$-AuNP-SWCNT model shows a resemblance to the corresponding theoretical model in number and placement of AuNPs along the SWCNT axis, while the $(GT)_6$-AuNP-SWCNT model shows clear regions of missing electron density. The initial N_{NP} was doubled for the starting $(GT)_6$-AuNP-SWCNT model (Figure 6-19C) leading to a slight reduction in χ^2-value and more consistent AuNPs (dummy-atom clusters) as shown in Figure 6-16D and Figure 6-24A. The best single SWCNT was again isolated from this revised double SWCNT model and compared against the corresponding theoretical model. The results were a clear improvement in number and placement of AuNPs along the SWCNT axis (Figure 6-24D). To confirm the *ab initio* modeling results, the two best starting models were switched for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs, resulting in a lower quality of fit in both cases (Figure 6-25) and supporting the ssDNA sequence-specific modeling outcomes.

It was realized that 1000 dummy atoms per AuNP was insufficient to properly model the ssDNA-AuNP-SWCNT systems resulting in the software compensating for lack of electron

density by expanding the model widths instead of lengths. Thus, to further explore whether the system could be modeled as a single SWCNT, the best fit double SWCNT model for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT was selected (Figure 6-24) and the single SWCNT region showing the clearest AuNP spacing was extracted (Figure 6-24C-D). Over many iterations, the positions and density (number of dummy atoms) of the AuNPs were carefully modified leading to a best fitting single SWCNT models (see SI Methods Section 6.5 and Figure 6-16A-C). The χ^2-value for the best single SWCNT model is far higher than the best double SWCNT model ($\chi^2 = 1.67$ vs 0.20). Upon closer inspection of the data, the residuals of the fit show that the single SWCNT models fit better at lowest q-values (Guinier region, defining the overall size) but worse in the midrange q-values (Porod region, defining the volume and morphology) than that of the double SWCNT model (Figure 6-16C). This observation suggests that the increased volume requirement may eventually be overcome with a sufficiently elongated and complex (more dummy atoms) model.

6.7 Chapter Supporting Information

6.7.1 Supplementary Figures and Tables
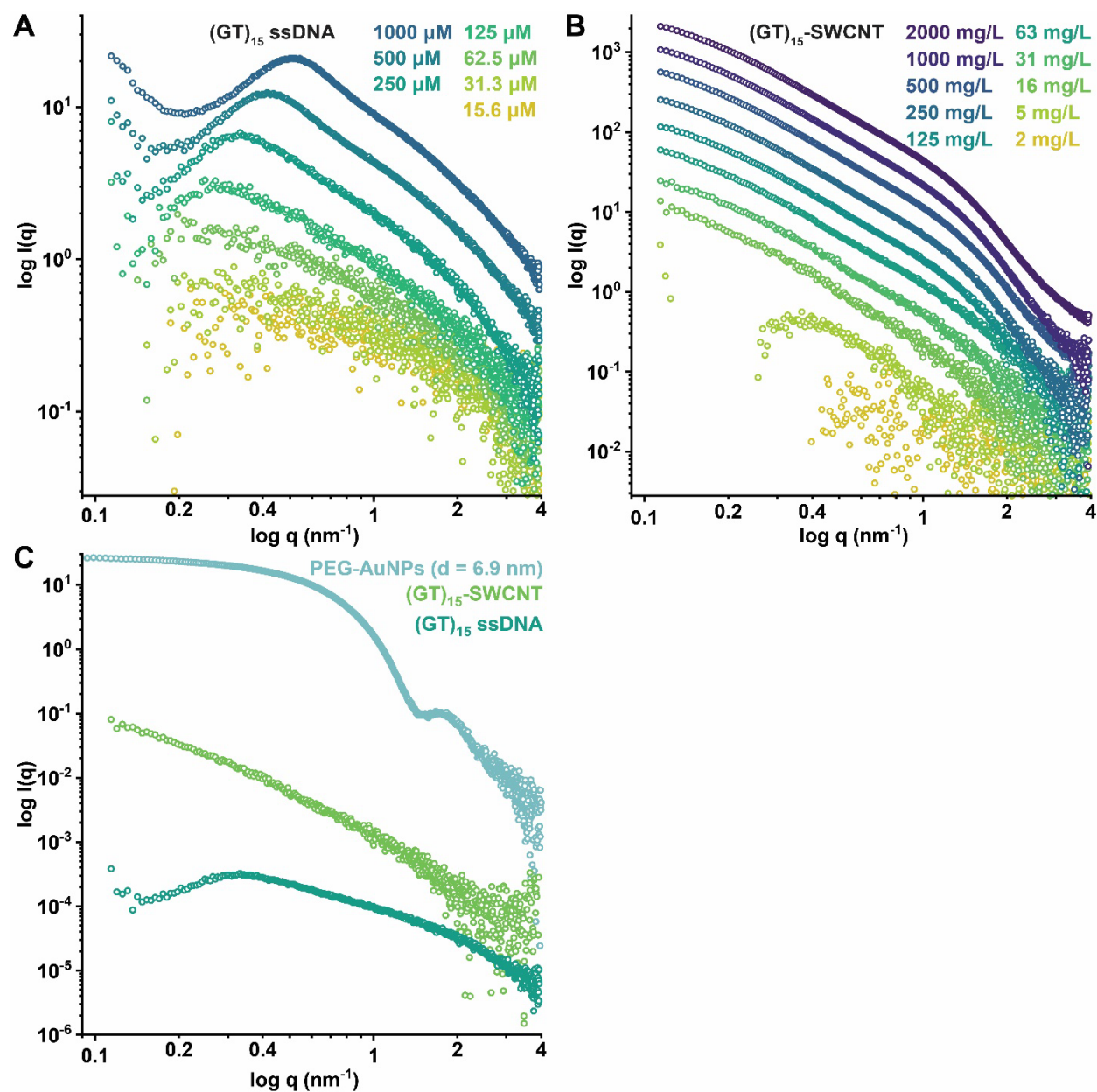
Figure 6-4. Scattering profiles as a function of concentration for serial dilutions of (A) $(GT)_{15}$ ssDNA alone and (B) $(GT)_{15}$-SWCNTs. (C) Absolute-scale scattering measurements for representative samples of 6.9 nm diameter PEG-AuNP, $(GT)_{15}$-SWCNTs, and $(GT)_{15}$ ssDNA alone, at the correct relative concentrations (i.e., 250 nM AuNP and ssDNA per 1 mg/L SWCNT).
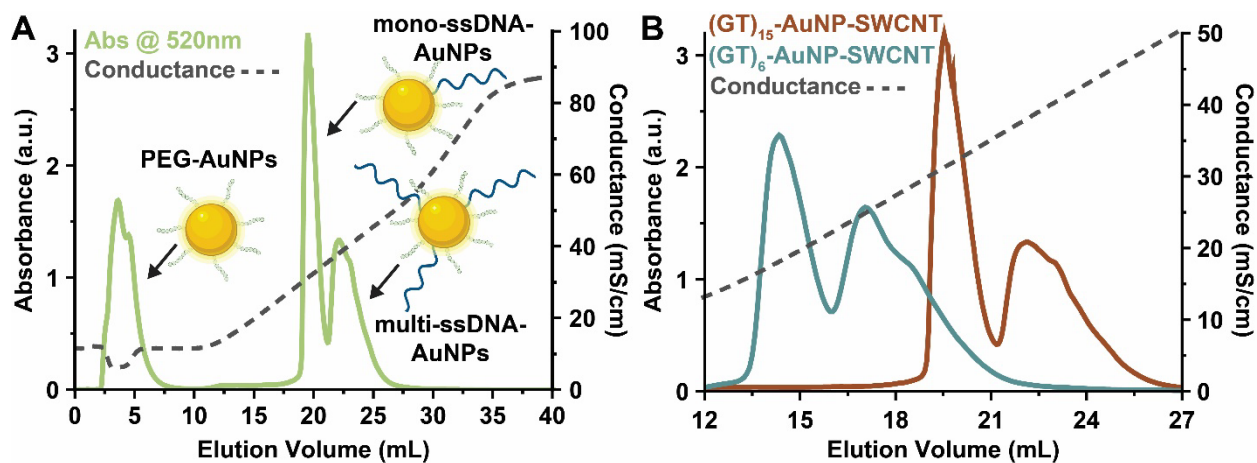
Figure 6-5. Anion exchange chromatograms for (A) the separation of non-conjugated PEG-AuNPs, mono-conjugated $(GT)_{15}$-AuNP, and multi-conjugated $(GT)_{15}$-AuNP using fast protein liquid chromatography (FPLC) and (B) $(GT)_{15}$- (red) vs. $(GT)_6$-AuNP (blue), demonstrating the length-dependent shift in the elution volume. Spectra are measured by diode array detector (DAD) at 520 nm. Conductance measurements depict salt gradient conditions (dash grey).
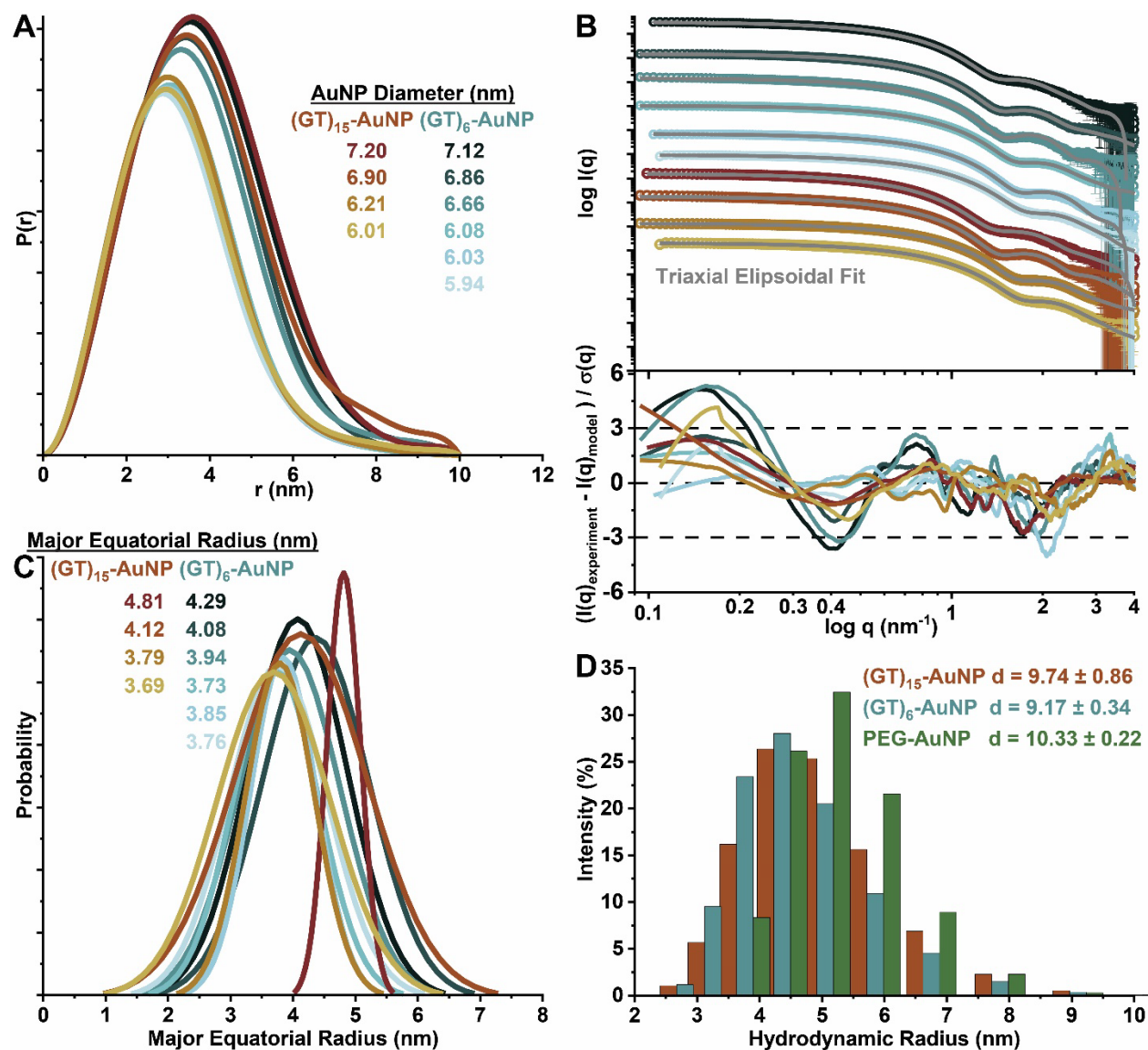
Figure 6-6. SAXS analysis for synthesized ssDNA-AuNPs. (A) Pairwise distribution functions, $P(r)$, for AuNPs of various sizes, scaled to the calculated average diameters for visual clarity. (B) Experimental SAXS profiles with calculated triaxial ellipsoidal fits (grey) for the prepared ssDNA-AuNPs (top) and the fit residuals (bottom). Scattering curves are offset for clarity and colored as in panel (A). Numerical values are summarized in Table 6-1 and Table 6-2. (C) Polydispersity of the major equatorial radius (r_A) modeled as a Gaussian distribution using SasView. Plots are scaled to the calculated r_A values and colored as in panel (A). Numerical values for panels (A-B) are summarized in Table 6-1 and Table 6-2. (D) DLS histograms reveal increased hydrodynamic radii of AuNPs after $(GT)_n$ ssDNA conjugation and mPEG-SH coating, as compared to diameters calculated from corresponding $P(r)$ functions (Table 6-3). Note that all ssDNA-AuNPs are also PEGylated.

Table 6-1. Physical parameters of synthesized $(GT)_6$-AuNPs obtained from pairwise distribution functions and triaxial ellipsoidal fits of scattering curves.

P(r) Peak Diameter (nm)	Major Equatorial Radius, r_A (nm)	Minor Equatorial Radius, r_B (nm)	Polar Radius, r_C (nm)	Polydispersity Index (PDI)	χ^2-value	r_A / r_C
5.94 ± 0.09	3.76 ± 0.08	2.88 ± 0.04	2.07 ± 0.01	0.21	0.74	1.8
6.03 ± 0.07	3.85 ± 0.03	2.86 ± 0.02	2.25 ± 0.01	0.14	2.48	1.7
6.08 ± 0.03	3.73 ± 0.16	3.08 ± 0.05	2.24 ± 0.04	0.18	0.88	1.7
6.66 ± 0.04	3.94 ± 0.12	3.26 ± 0.18	2.61 ± 0.16	0.2	2.41	1.5
6.86 ± 0.04	4.08 ± 0.03	3.43 ± 0.01	2.60 ± 0.01	0.19	0.66	1.6
7.12 ± 0.06	4.29 ± 0.04	3.65 ± 0.02	2.60 ± 0.01	0.2	1.96	1.7

Table 6-2. Physical parameters of synthesized $(GT)_{15}$-AuNPs obtained from pairwise distribution functions and triaxial ellipsoidal fits of scattering curves.

P(r) Peak Diameter (nm)	Major Equatorial Radius, r_A (nm)	Minor Equatorial Radius, r_B (nm)	Polar Radius, r_C (nm)	Polydispersity Index (PDI)	χ^2-value	r_A / r_C
6.01 ± 0.08	3.69 ± 0.08	2.97 ± 0.03	2.15 ± 0.01	0.25	1.21	1.7
6.21 ± 0.08	3.79 ± 0.12	2.97 ± 0.06	2.20 ± 0.01	0.15	0.7	1.7
6.90 ± 0.02	4.12 ± 0.11	3.39 ± 0.03	2.65 ± 0.03	0.26	0.33	1.6
7.20 ± 0.02	4.81 ± 0.15	3.48 ± 0.09	2.64 ± 0.02	0.06	1.1	1.8

Table 6-3. Physical parameters of synthesized AuNPs as obtained from pairwise distribution functions of X-ray scattering and dynamic light scattering.

	P(r) Peak Diameter (nm)	DLS Peak Diameter (nm)	Increase in Diameter from PEGylation (nm)
$(GT)_{15}$-AuNP	6.90 ± 0.02	9.74 ± 0.86	2.83 ± 0.88
$(GT)_6$-AuNP	6.86 ± 0.04	9.17 ± 0.34	2.31 ± 0.38
PEG-AuNP	6.82 ± 0.03	10.33 ± 0.22	3.51 ± 0.25
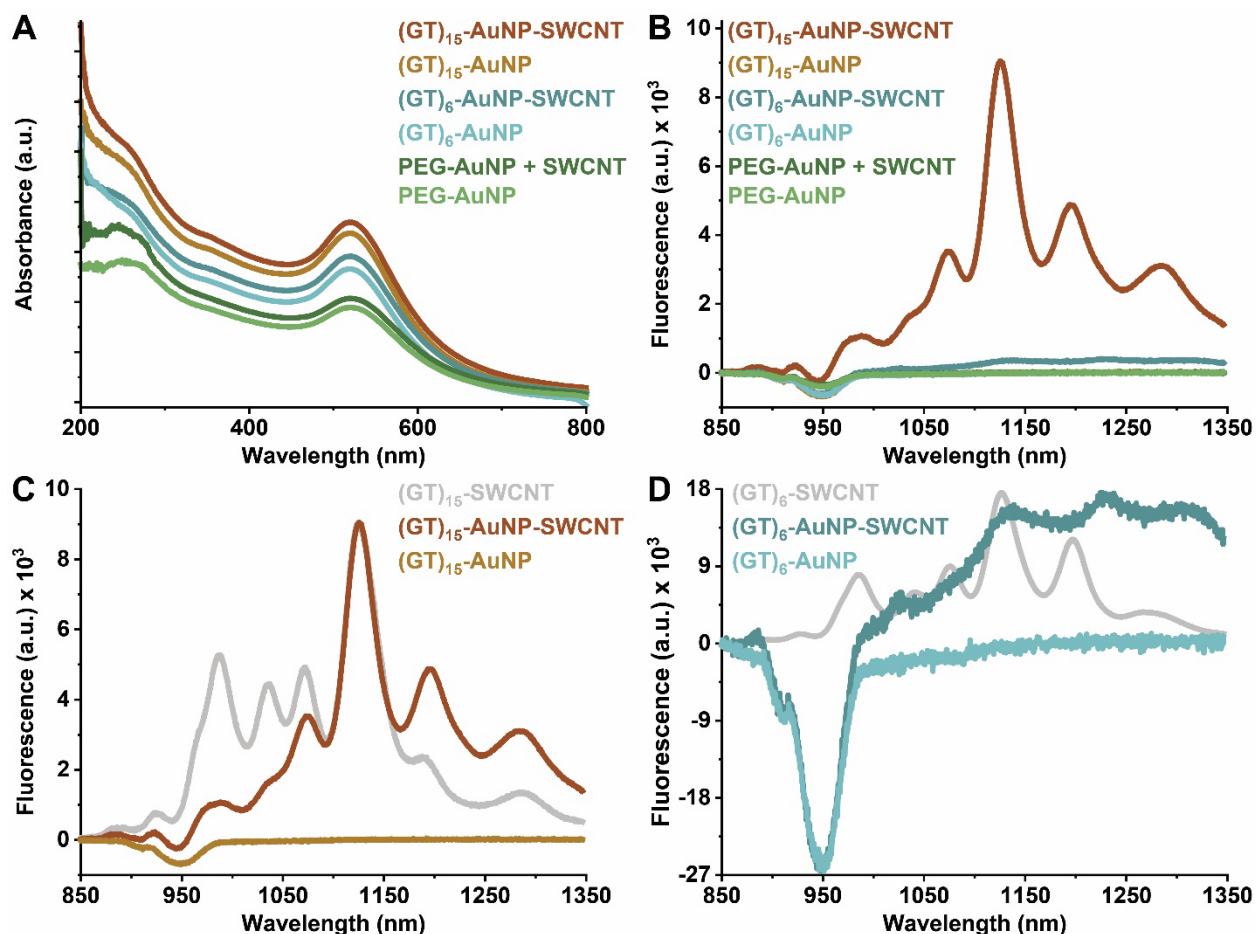
Figure 6-7. ssDNA-AuNP-SWCNT optical characterization. (A) Absorbance spectra for $(GT)_{15}$-AuNP alone and suspending SWCNTs (orange and red, respectively), $(GT)_6$-AuNP alone and suspending SWCNTs (light and dark blue, respectively), and PEG-AuNP alone and attempted-to-suspend SWCNTs (light and dark green, respectively). The consistent AuNP plasmon resonance peak at approximately 520 nm demonstrates that AuNP tags are intact after the probe-tip sonication suspension process with SWCNTs. Spectra are offset for clarity. Note that these SWCNT concentrations of approximately 0.2 mg/L produce negligible near-infrared absorbance fingerprints and this region of the spectrum is therefore omitted. (B) Fluorescence spectra for the same sample set as panel (A) confirm SWCNT suspension with the presence of near-infrared fluorescence. (C) $(GT)_{15}$-AuNP-SWCNT and $(GT)_{15}$-SWCNT samples are compared to a $(GT)_{15}$-SWCNT conjugate without AuNP tags, with peaks normalized to maximum emission intensity. (D) $(GT)_6$-AuNP-SWCNT and $(GT)_6$-SWCNT samples are compared to a $(GT)_6$-SWCNT conjugate without AuNP tags, with peaks normalized to maximum emission intensity. All AuNPs are synthesized diameters of 5.9-7.2 nm. All fluorescence measurements were obtained with 721 nm laser excitation.
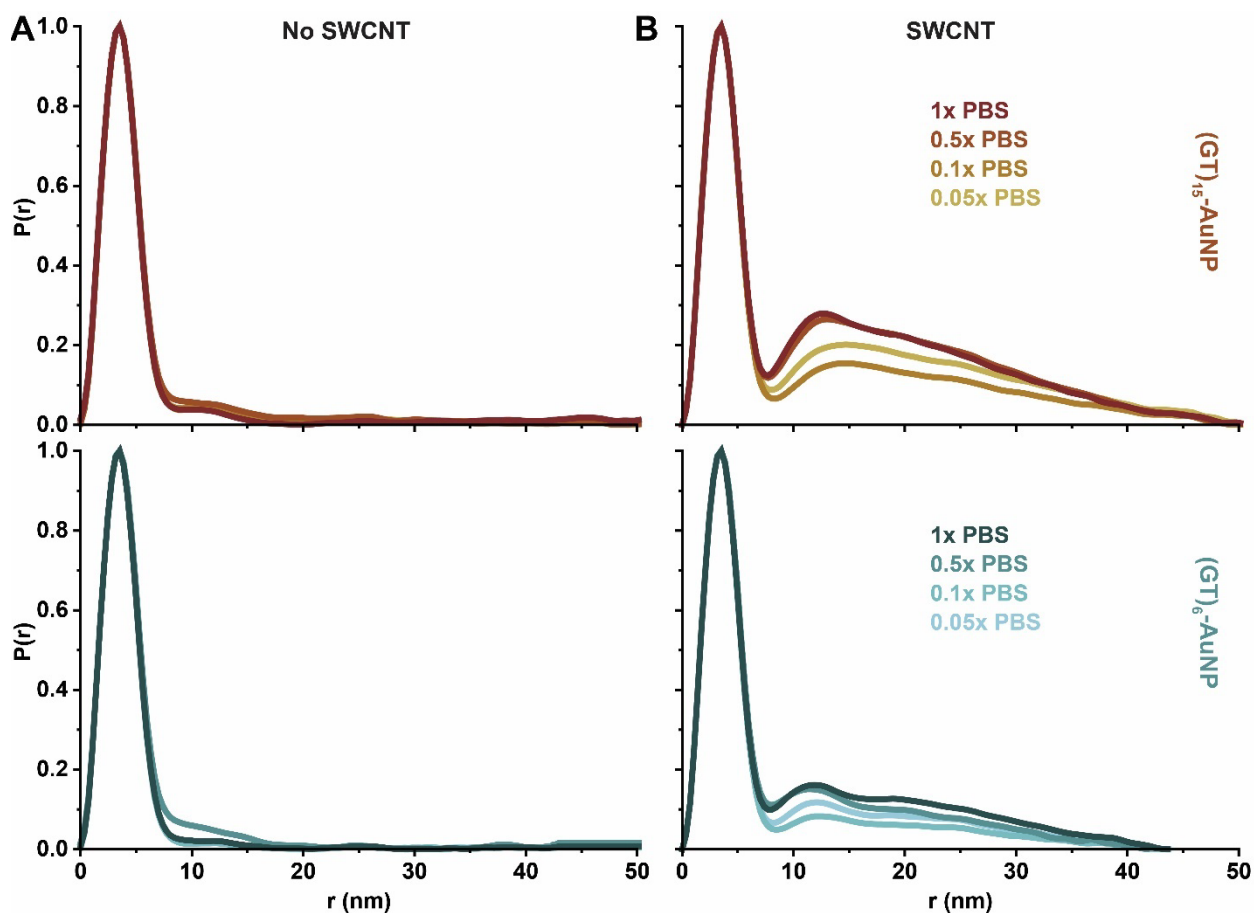
Figure 6-8. Representative pairwise distribution functions, *P(r)*, of AuNP-tagged ssDNA either (A) free in solution or (B) adsorbed to SWCNTs, as a function of ionic strength over a range of 0.05X to 2X PBS or corresponding Debye lengths, λ_D = 3.37-0.53 nm. (Top) (GT)$_{15}$-AuNPs vs. (GT)$_{15}$-AuNP-SWCNT complexes and (bottom) (GT)$_6$-AuNPs vs. (GT)$_6$-AuNP-SWCNT complexes. *P(r)* functions are normalized to the intra-AuNP peak to compensate for slight fluctuations in X-ray beam intensity or sample concentration. Order only emerges in the presence of ssDNA-AuNPs adsorbed to SWCNTs.
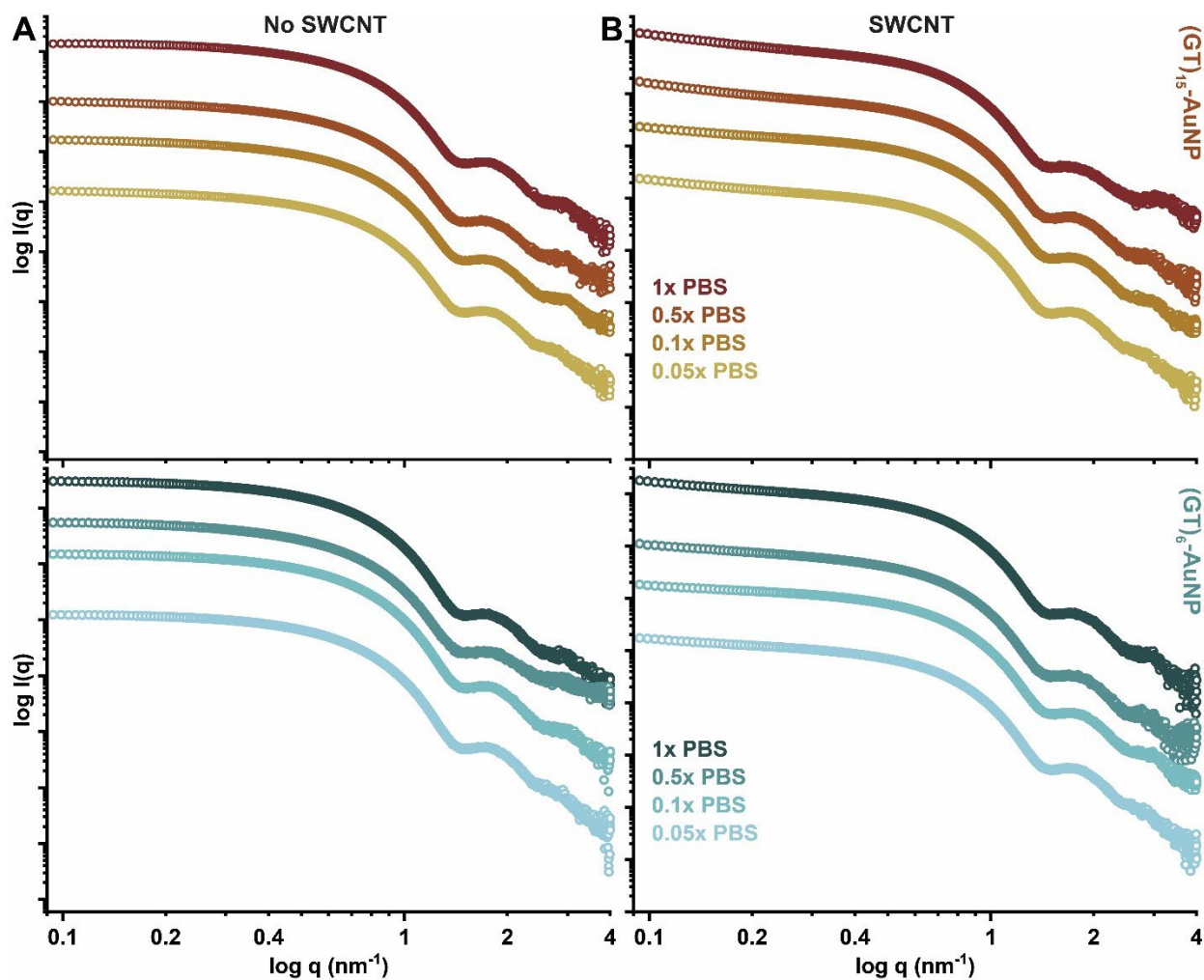
Figure 6-9. Representative scattering curves of AuNP-tagged ssDNA either (A) free in solution or (B) adsorbed to SWCNTs, as a function of ionic strength over a range of 0.05X to 2X PBS or corresponding Debye lengths, λ_D = 3.37-0.53 nm. (Top) $(GT)_{15}$-AuNPs vs. $(GT)_{15}$-AuNP-SWCNT complexes and (bottom) $(GT)_6$-AuNPs vs. $(GT)_6$-AuNP-SWCNT complexes. Scattering curves are offset for clarity. The major difference of note in curves (A) without vs. (B) with SWCNTs is found in the lowest q-value range (q < 0.2 nm^{-1}): the upward curvature at low q with SWCNTs represents interparticle interaction from the ordering of AuNPs on the SWCNT surface.
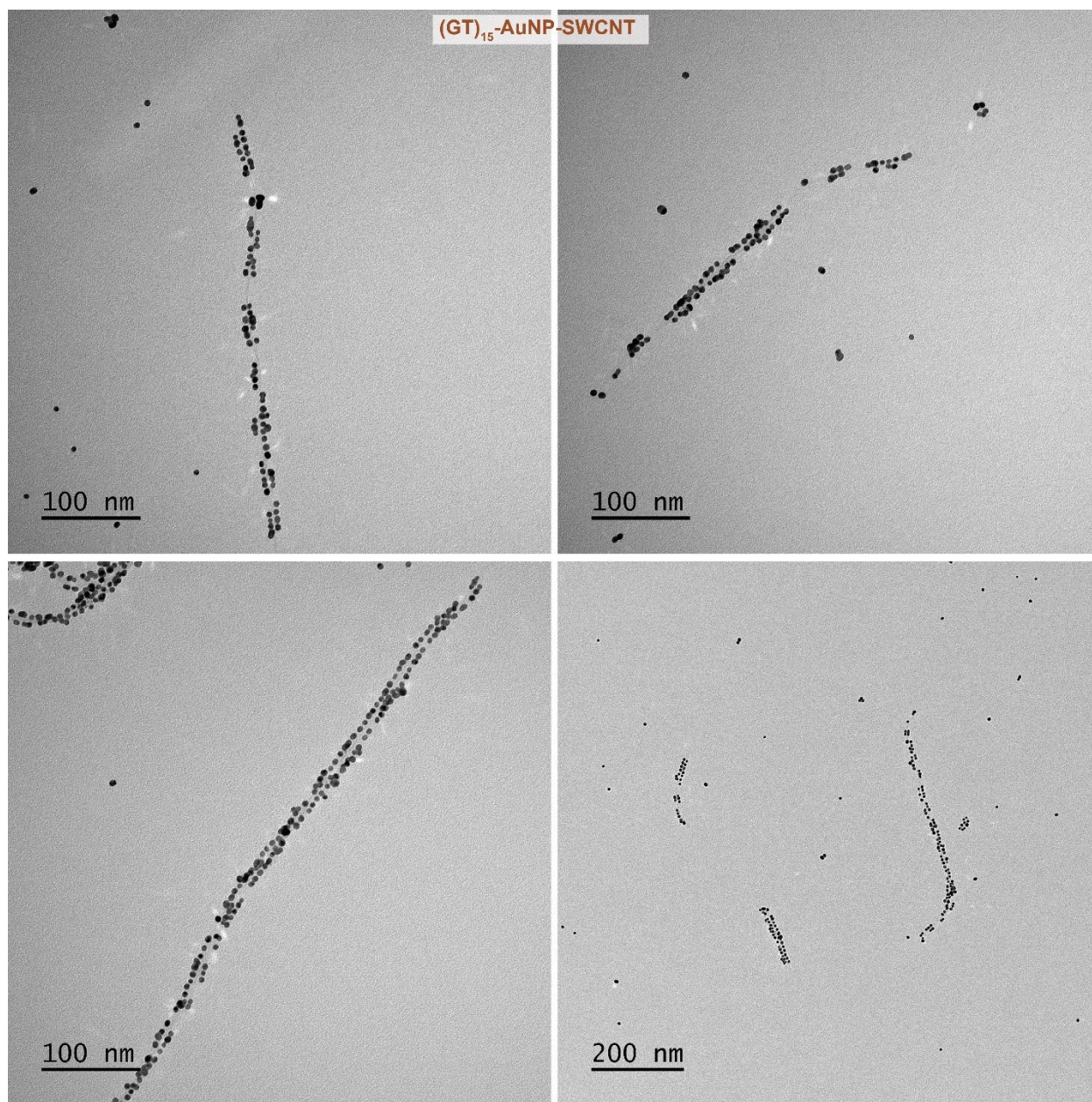
Figure 6-10. Representative TEM images for (GT)$_{15}$-AuNP-SWCNTs.
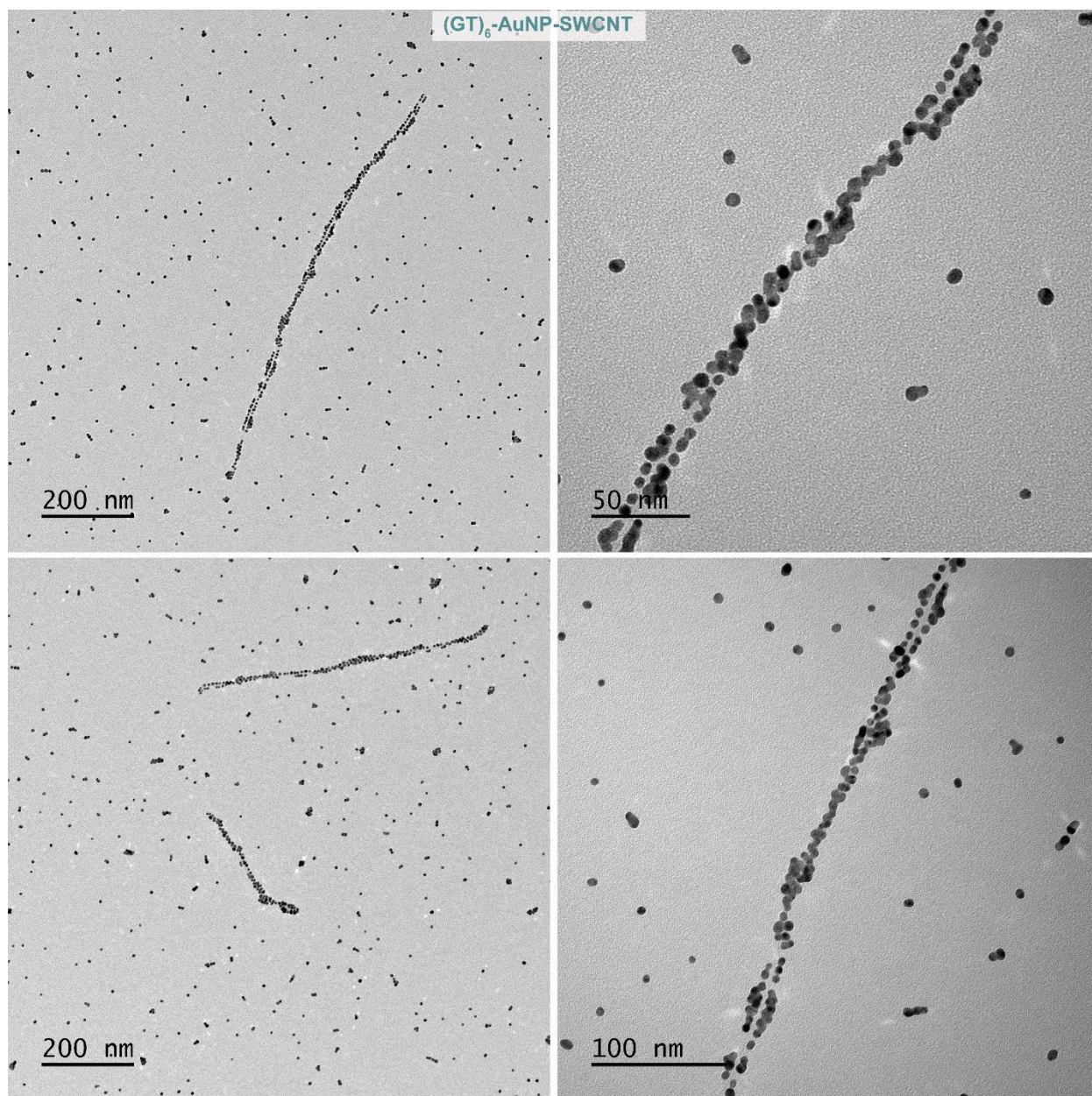
Figure 6-11. Representative TEM images for (GT)$_6$-AuNP-SWCNTs.
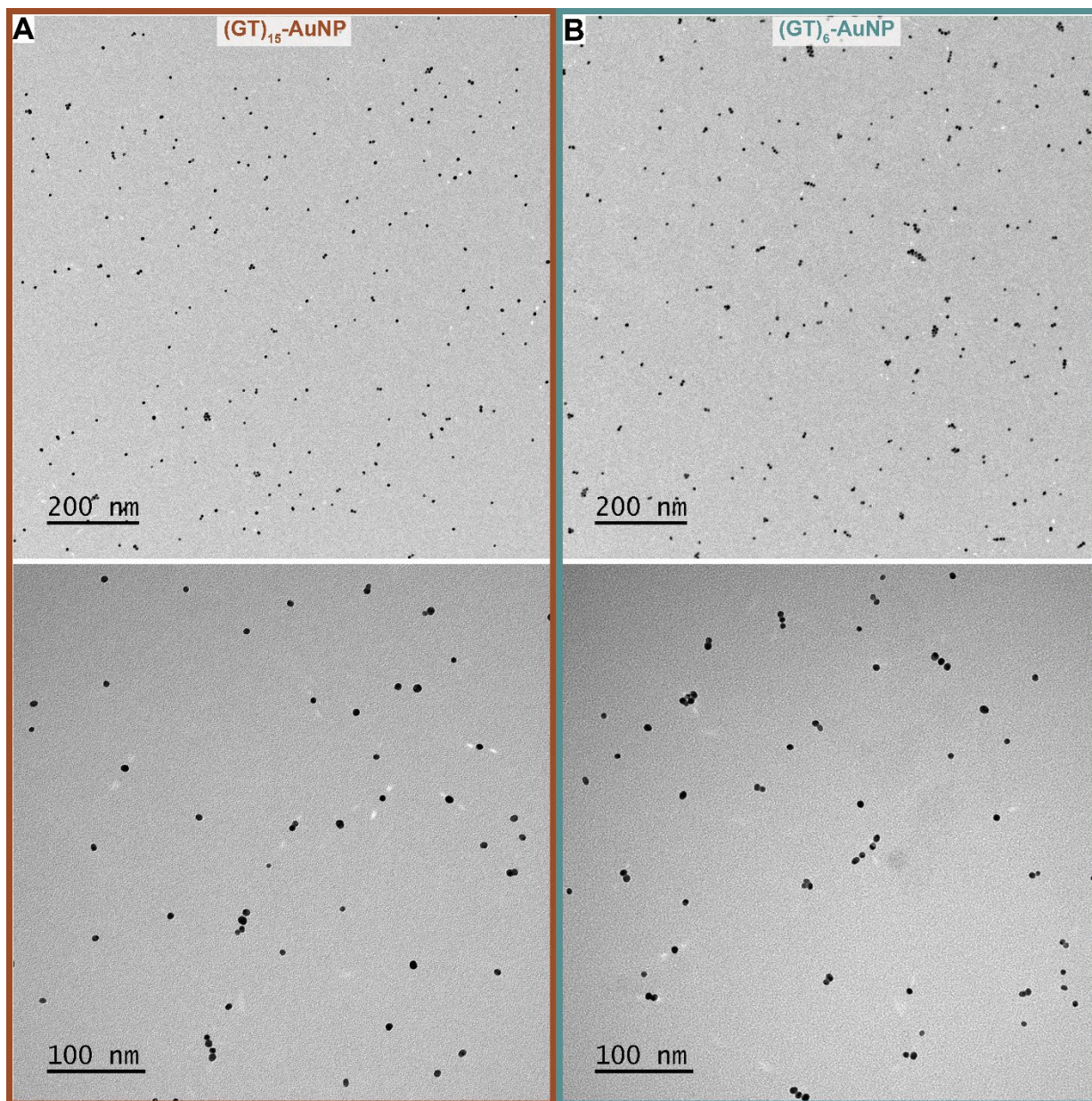
Figure 6-12. Representative TEM images for (A) $(GT)_{15}$-AuNPs and (B) $(GT)_6$-AuNPs prepared by the same method as ssDNA-AuNP-SWCNTS but absent the SWCNT substrate. All controls do not show order when free in the solution state, in the absence of SWCNTs.
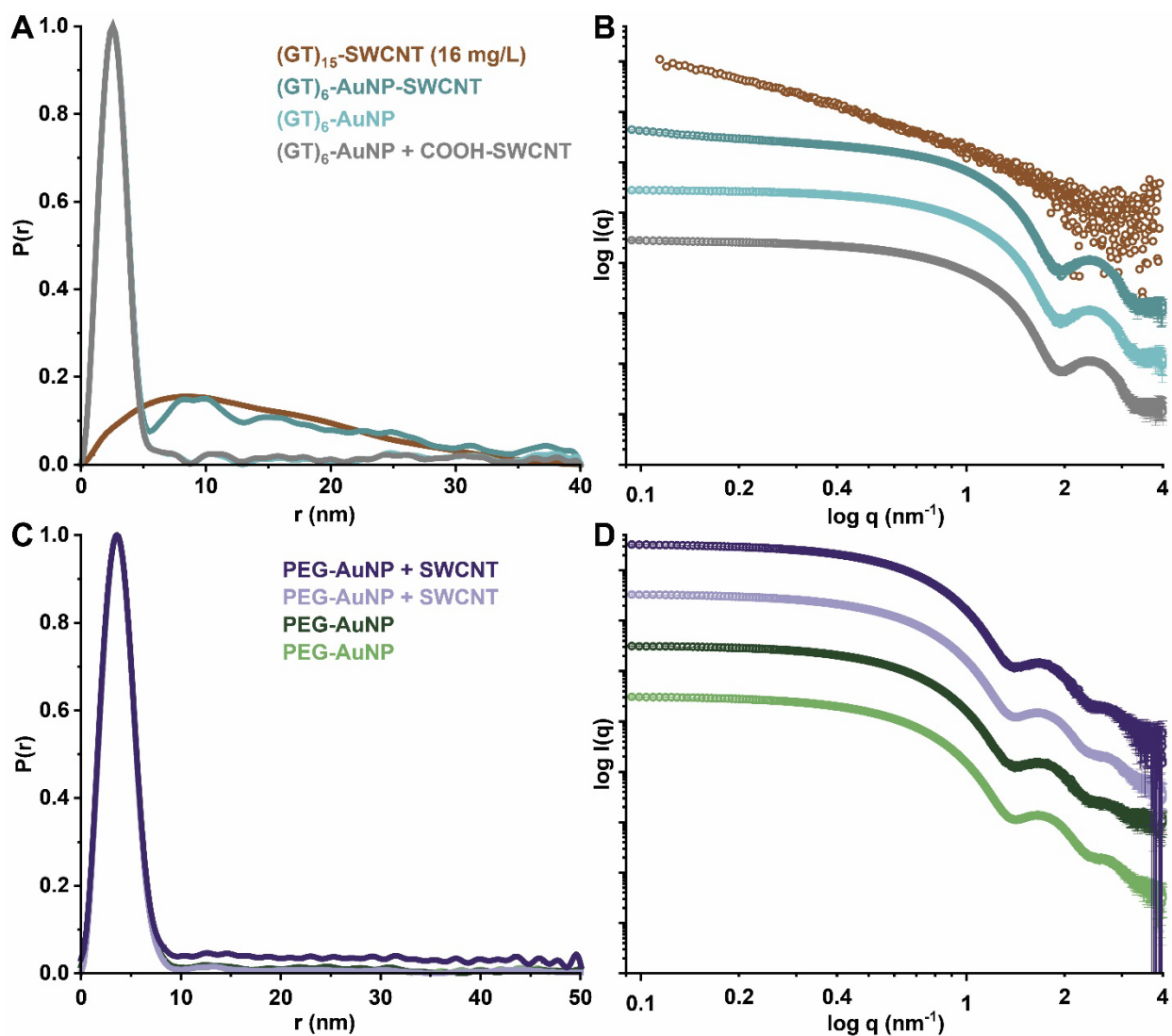
Figure 6-13. Experimental controls for ssDNA-AuNP-SWCNT preparations. (A) Pairwise distribution functions, $P(r)$, and (B) scattering curves for $(GT)_{15}$-SWCNTs (no AuNPs) at the lowest measurable concentration (16 mg/L), $(GT)_6$-AuNP-SWCNTs (full complex), $(GT)_6$-AuNPs (no SWCNTs), and $(GT)_6$-AuNP-SWCNTs (2 mg/L), and $(GT)_6$-AuNPs mixed with carboxylated SWCNTs (2 mg/L; no suspension). The $P(r)$ function for $(GT)_{15}$-SWCNT (no AuNPs) is normalized to the inter-AuNP peak of $(GT)_6$-AuNPs-SWCNT for clarity. (C) $P(r)$ function and (D) scattering curves for two batches of PEG-AuNPs (green) vs. PEG-AuNPs attempted-to-suspend with SWCNTs (purple) by the same method as ssDNA-AuNP-SWCNTS. All AuNP samples are normalized to intra-AuNP peak.
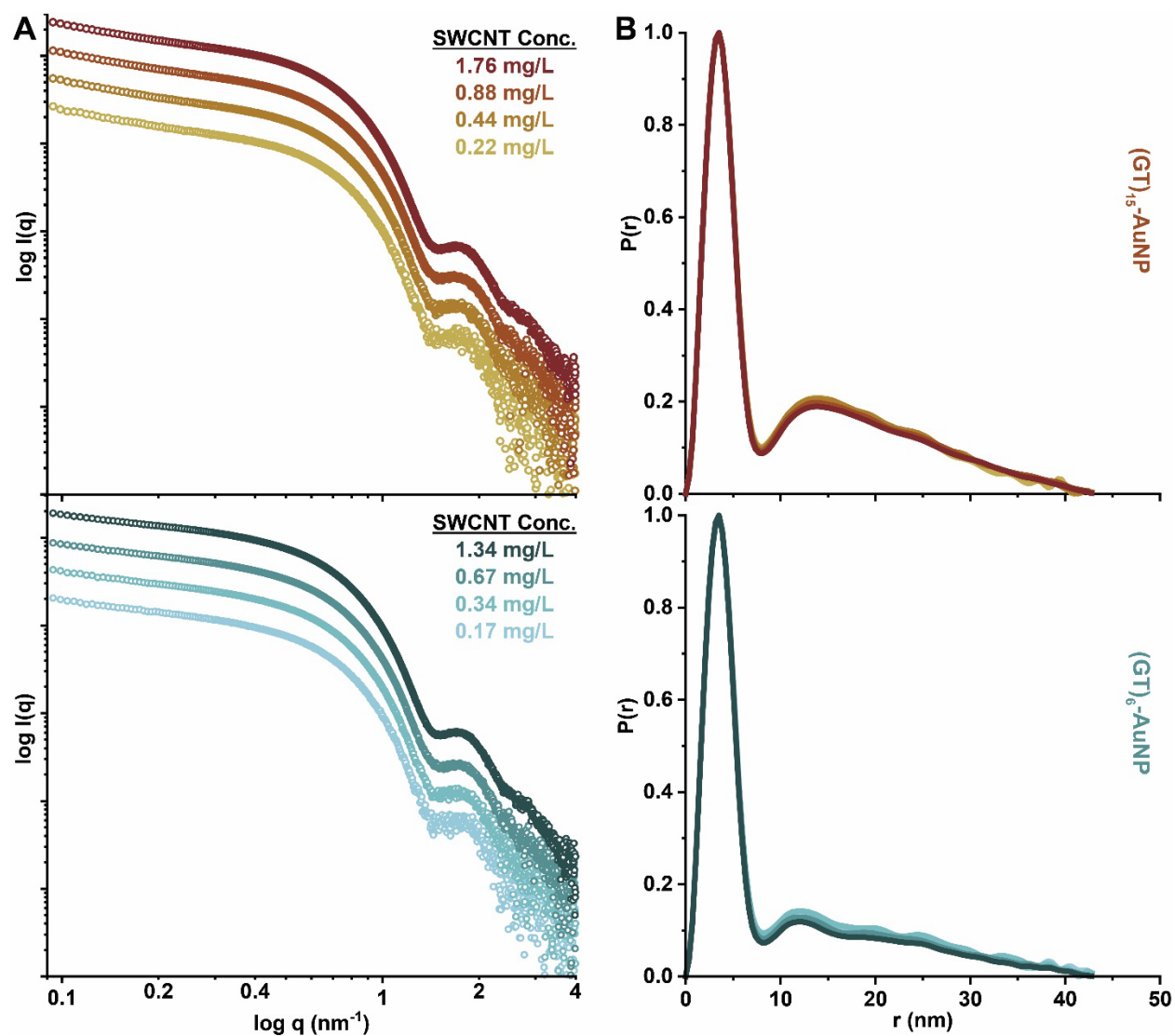
Figure 6-14. (A) Scattering curves of AuNP-tagged ssDNA adsorbed to SWCNTs, as a function of SWCNT concentration for (top) $(GT)_{15}$-AuNP-SWCNT complexes and (bottom) $(GT)_6$-AuNP-SWCNT complexes. Scattering curves are offset for clarity. (B) Pairwise distribution functions, *P(r)*, of AuNP-tagged ssDNA adsorbed to SWCNTs, as a function of concentration. *P(r)* functions are normalized to the intra-AuNP peak to compensate for fluctuations in intensity. Samples are observed over SWCNT concentrations relevant to biological applications (0.17-1.76 mg/L).
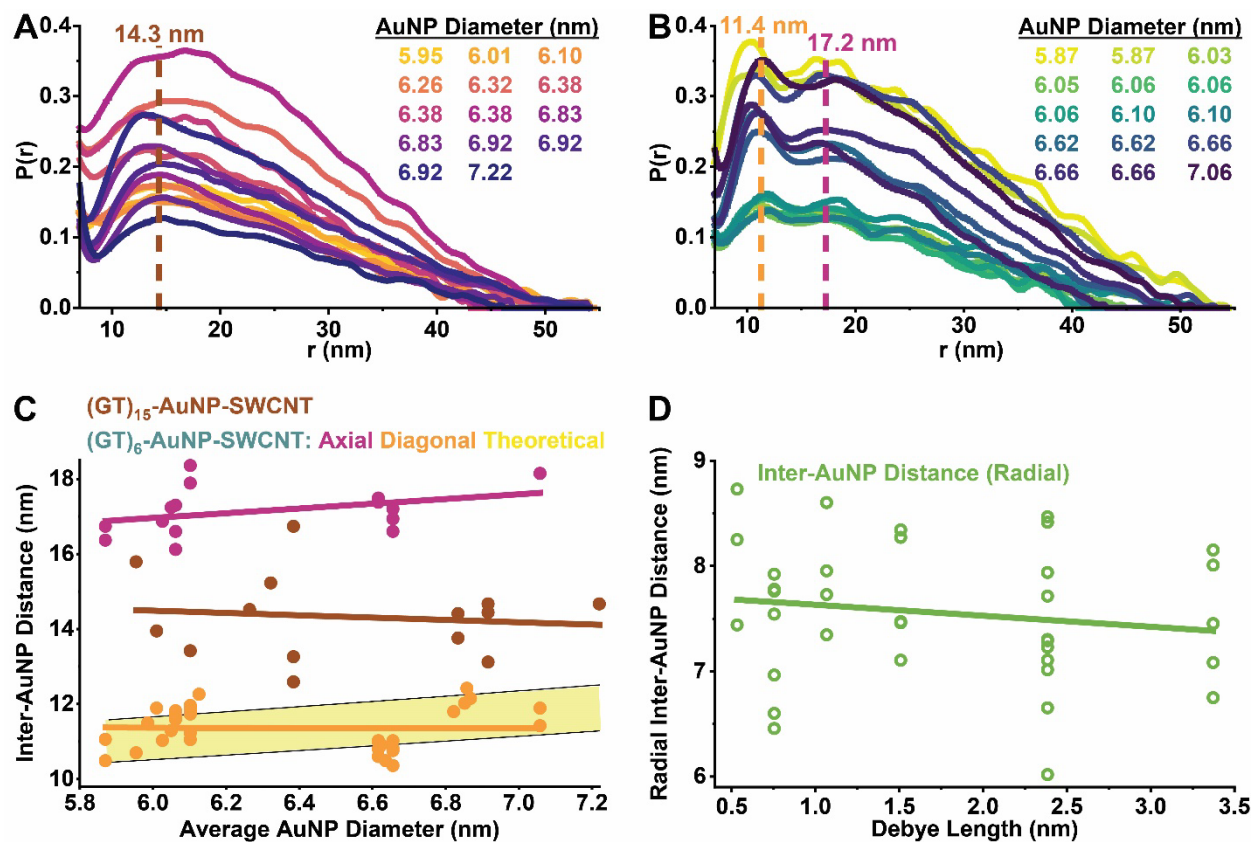
Figure 6-15. Pairwise distribution functions, *P(r)*, for samples with varying AuNP sizes ranging from 5.9 to 7.2 nm diameter for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs that demonstrate a clear second inter-AuNP distance. Dashed vertical lines are added to visualize the average inter-AuNP distances for (GT)$_{15}$-AuNP-SWCNTs (red), and (GT)$_6$-AuNP-SWCNTs diagonal (orange) and axial (magenta). *P(r)* functions are normalized to the primary intra-AuNP peak, then the x-axis minimum is set to focus on the inter-AuNP peak for clarity. (C) Plots inter-AuNP distances for (GT)$_{15}$-AuNP-SWCNTs (red), and (GT)$_6$-AuNP-SWCNTs diagonal (orange) and axial (magenta) as a function of AuNP size. Only the diagonal inter-AuNP distances should be subject to changes as a function of AuNP size as displayed graphically by theoretical distances (yellow) calculated with AuNPs fixed on the SWCNT surface. (D) Plots of radial inter-AuNP distances (green) for (GT)$_6$-AuNP-SWCNTs as a function of Debye length (λ_D) as calculated from experimental diagonal and axial inter-AuNP distances from a single size of AuNPs (d = 6.1 ± 0.03 nm).
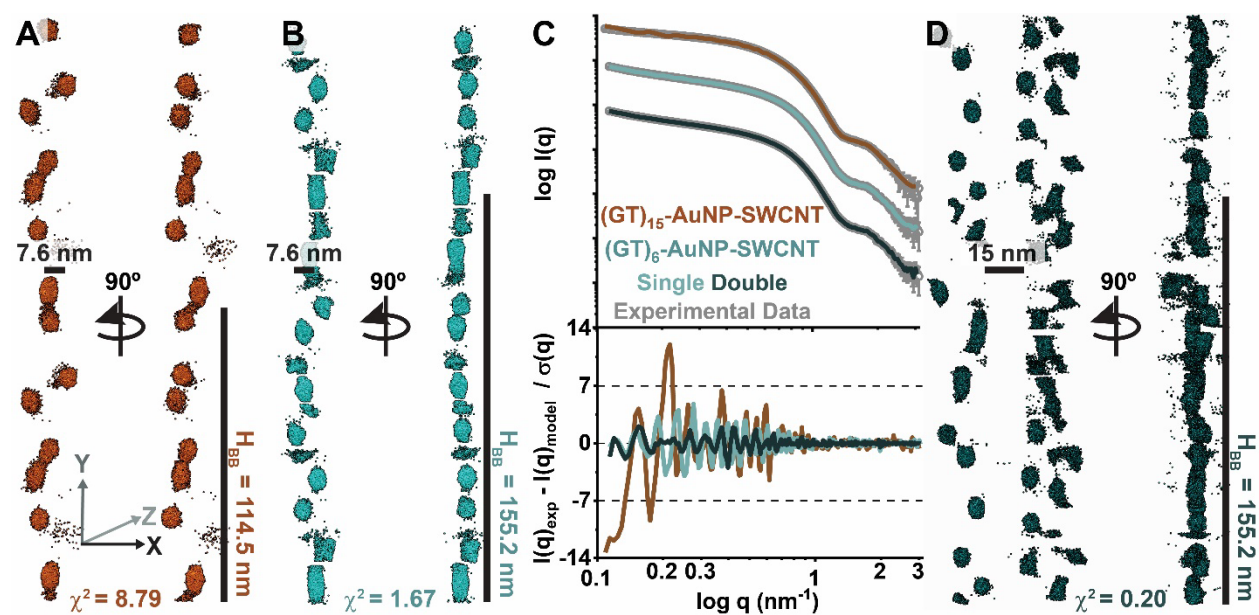
Figure 6-16. Best fit single SWCNT *ab initio* modeling results for (A) $(GT)_{15}$-AuNP-SWCNT and (B) $(GT)_6$-AuNP-SWCNTs. (C) SAXS profiles with fits and residuals for each model colored as in panel (A, B, and D). Scattering curves are offset for clarity. (D) *Ab initio* modeling results for $(GT)_6$-AuNP-SWCNTs modeled as two parallel SWCNTs for comparison. Initial models started with a stack building block height (H_{BB}) of 114.5 and 155.2 nm for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT, respectively, as defined by the best fit starting model parameters found in Table 4-1. Final χ^2-values shown beneath each model.
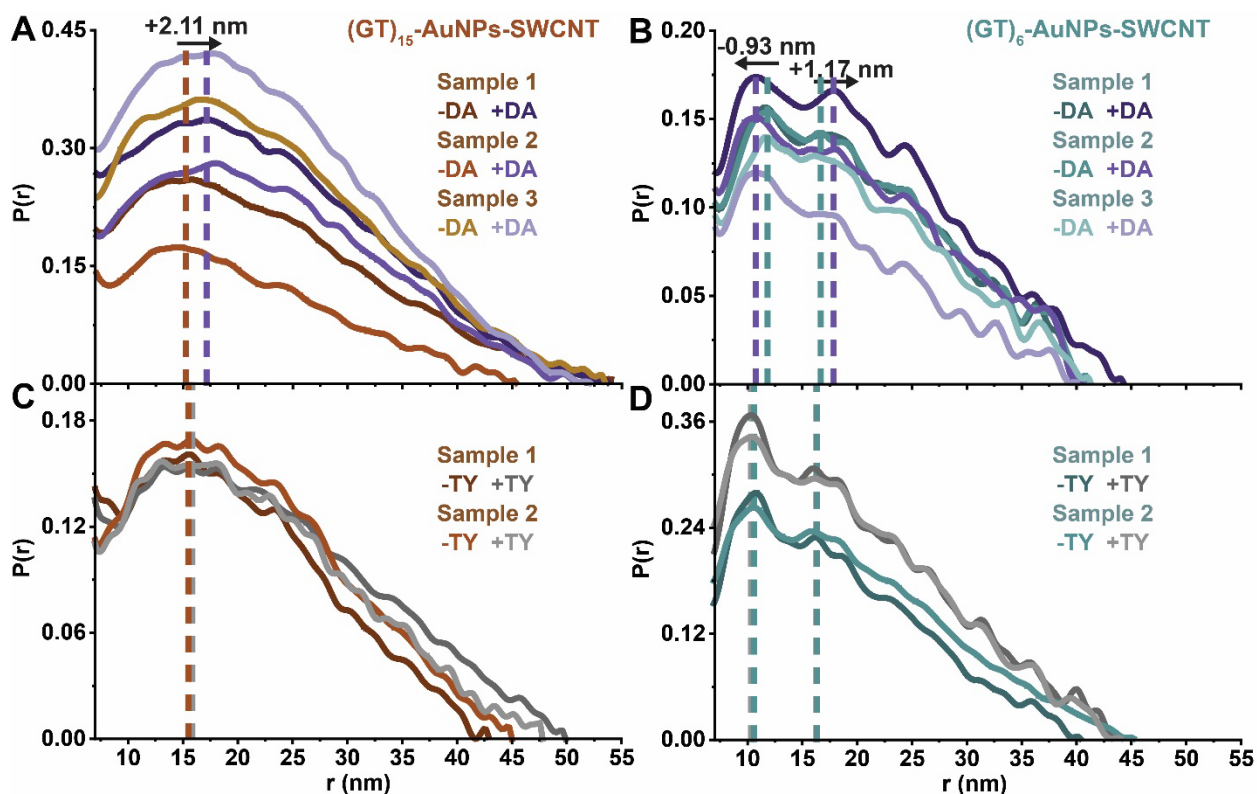
Figure 6-17. Pairwise distribution functions, *P(r)*, for all replicates alone for (A-C) (GT)$_{15}$-AuNP-SWCNTs (red-orange series) and (B-D) (GT)$_6$-AuNP-SWCNTs (blue series) or in the presence of (A-B) DA (purple series) or (C-D) TY (grey series). Dashed vertical lines are added to visualize peak shifts. *P(r)* functions are normalized to the primary intra-AuNP peak, then the x-axis minimum is set to focus on the inter-AuNP peak for clarity.
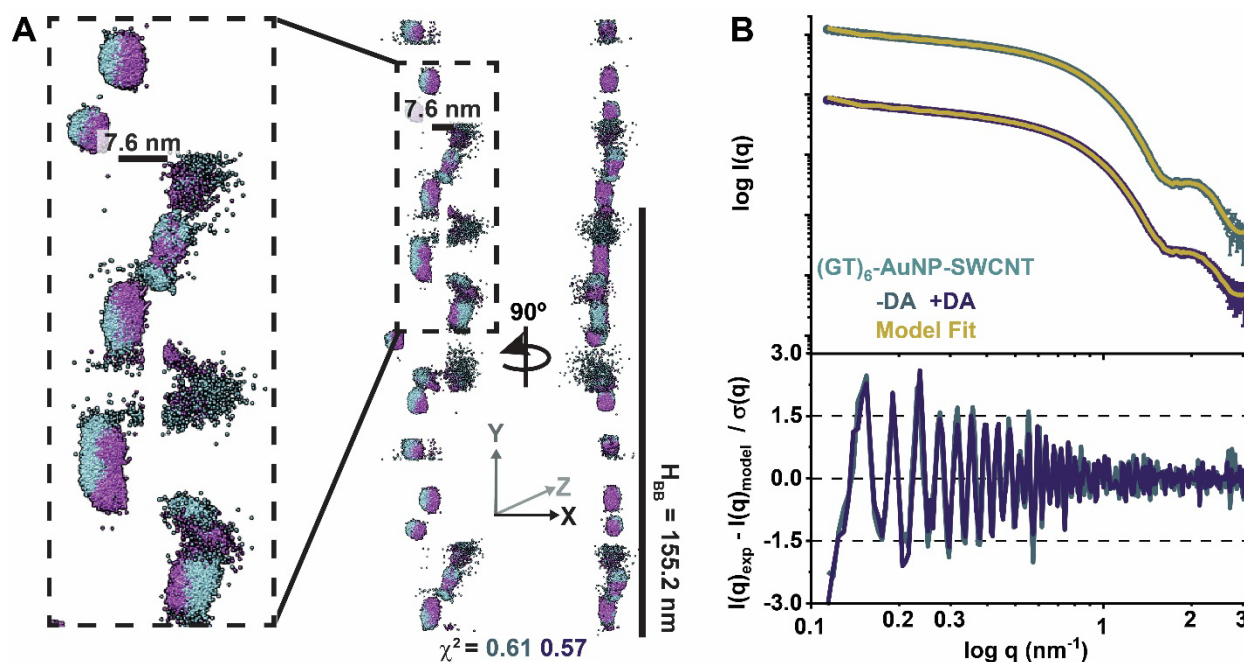
Figure 6-18. (A) Comparison of *Ab initio* modeling results for $(GT)_6$-AuNP-SWCNTs modeled from scattering profiles with (blue series) and without dopamine (DA; purple series). Initial model started from the best single SWCNT model (Figure 6-1F and Figure 6-16C) started with two parallel rows of AuNPs set 7.5 nm apart. Both models start with a stack building block height (H_{BB}) of 155.2 nm and number of stacks (N_S) of 2 as defined by the best fit starting model parameters found in Table 4-1. Final χ^2-values shown beneath each model. (B) SAXS profiles with model fits and residuals for each complex colored as in panel (A). Scattering curves are offset for clarity.

Figure 6-19. Schematic demonstration of adapted SASHEL *ab initio* modeling strategy with labeling of important definitions for stack building block heights (H_{BB}), inter-AuNP block heights (H_{GB}), number of stacks (N_S), and number of AuNPs (N_{NP}). The stepwise methodology is as follows: (A) broad movement of 2000 initial dummy atoms from initial guess core-shell model (green) resulting in a densely packed strand of dummy-atom clusters, each representative of an AuNP (pink). (i) One clearly defined cluster is selected and duplicated to give an AuNP pair, (ii) then the pair of AuNPs is expanded to 1000 dummy atoms per AuNP (purple). Each AuNP in the pair is maneuvered into desired initial geometries as determined experimentally, representing one H_{GB} unit, unique for (B) $(GT)_{15}$-AuNP-SWCNTs and (C) $(GT)_6$-AuNP-SWCNTs. (iii) The H_{GB} is replicated in one direction using symmetrical boundary conditions an integer (k) number of times to produce a total stack height, $H_{BB} = k*H_{GB}$, using parameters in Table 4-1. (iv) The starting model is then used to produce ab initio models representing two neighboring ssDNA-AuNP-SWCNTs. (v) The single ssDNA-AuNP-SWCNT showing the clearest AuNP spacing is selected and the regions of electron density are replaced with denser 3000 dummy-atom clusters (AuNPs).

145

Table 6-4. AuNP starting strand distances series parameters and fitting results for SASHEL modeling.

	Starting AuNP Strand Distances (nm)	10	15	20
χ^2-value of Fit	(GT)$_{15}$-AuNP-SWCNT	2.10	0.27	0.99
	(GT)$_6$-AuNP-SWCNT	6.72	1.19	1.35

Table 6-5. Stack parameters for SASHEL modeling.

	Number of Stacks (N_S)	16	8	5	4	3	3	2	2	2
	AuNPs per Stack (N_{NP})	2	4	6	8	10	12	14	16	18
Total Stack Height (nm)	(GT)$_{15}$-AuNP-SWCNT	14.3	28.6	42.9	57.2	71.6	85.9	100.2	114.5	128.8
	(GT)$_6$-AuNP-SWCNT	17.2	34.5	51.7	69.0	86.2	103.5	120.7	137.9	155.2
χ^2-value of Fit	(GT)$_{15}$-AuNP-SWCNT	284.2	14.8	3.11	0.51	0.70	3.37	0.97	0.27	0.41
	(GT)$_6$-AuNP-SWCNT	99.8	9.77	1.19	4.60	1.05	0.34	0.96	0.48	0.24
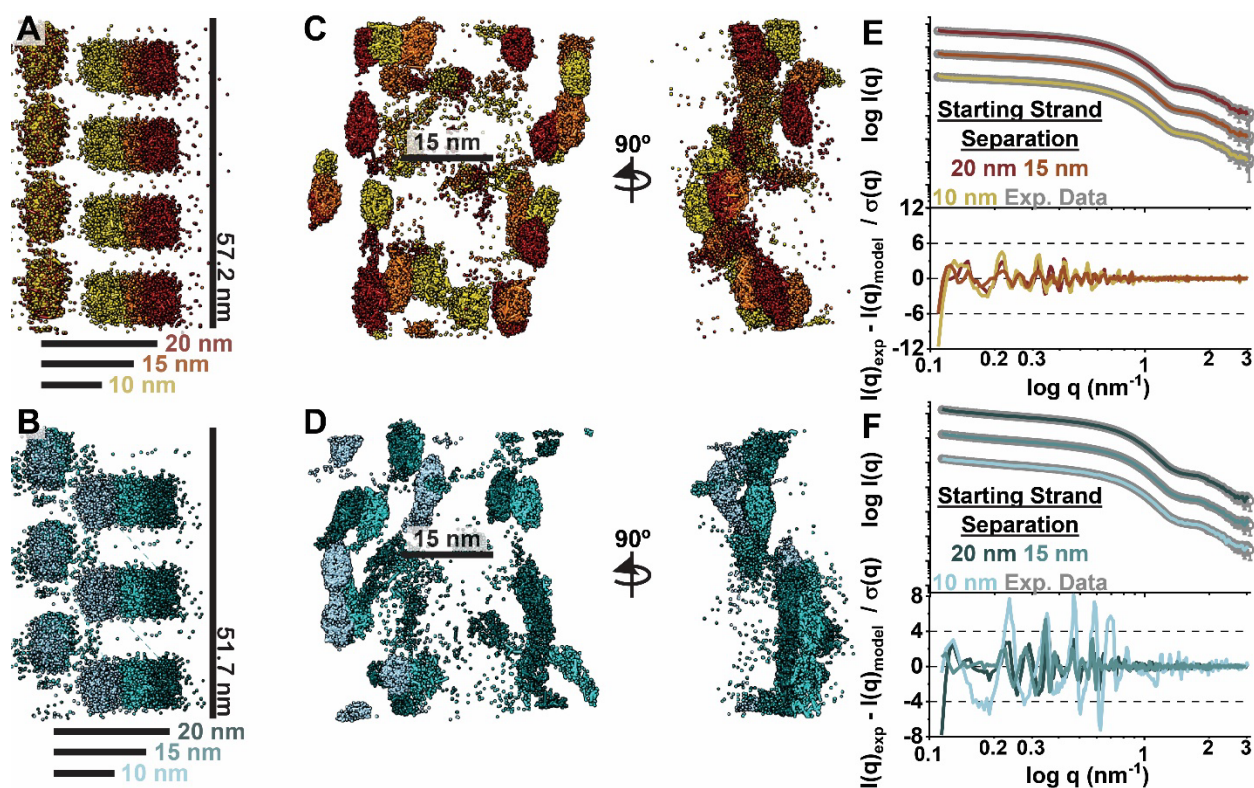
Figure 6-20. *Ab initio* modeling series for different starting AuNP-tagged ssDNA strand distances of 10, 15, and 20 nm via SASHEL, with full parameters shown in Table 6-4. Starting models for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs. *Ab initio* modeling results for (C) (GT)$_{15}$-AuNP-SWCNTs and (D) (GT)$_6$-AuNP-SWCNTs. SAXS profiles with model fits and residuals for (E) (GT)$_{15}$-AuNP-SWCNTs and (F) (GT)$_6$-AuNP-SWCNTs with each complex colored as in panel (A) and (B), respectively. Scattering curves are offset for clarity.
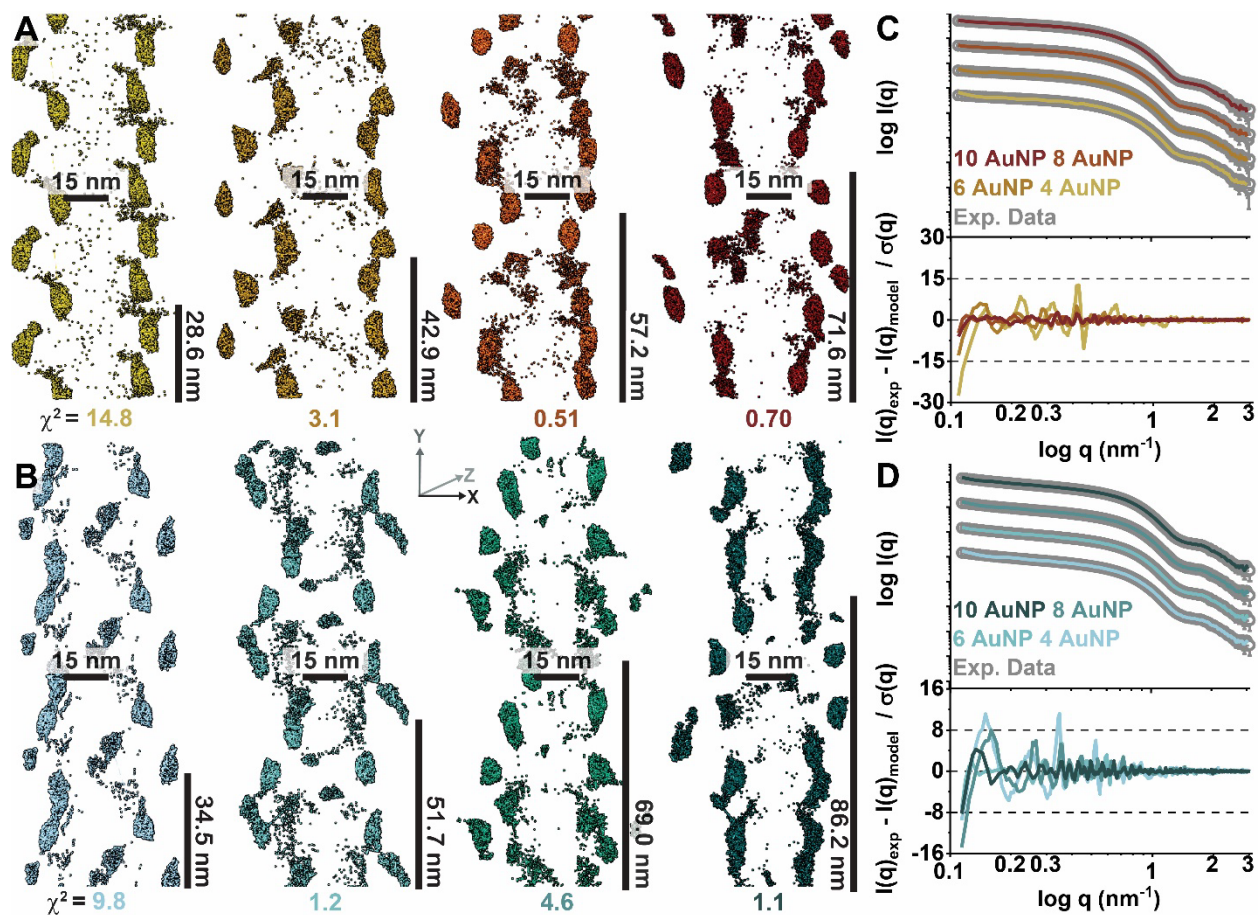
Figure 6-21. *Ab initio* modeling series results for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs made via SASHEL for 4 to 10 AuNPs per inter-AuNP block height (H_{GB}), as defined in SI Methods Section 6.5 and Figure 6-19. χ^2-values are shown below each model and starting parameters are shown in Table 4-1. SAXS profiles with model fits and residuals for (C) (GT)$_{15}$-AuNP-SWCNTs and (D) (GT)$_6$-AuNP-SWCNTs, with each sample colored as in panel (A) and (B), respectively. Scattering curves are offset for clarity.
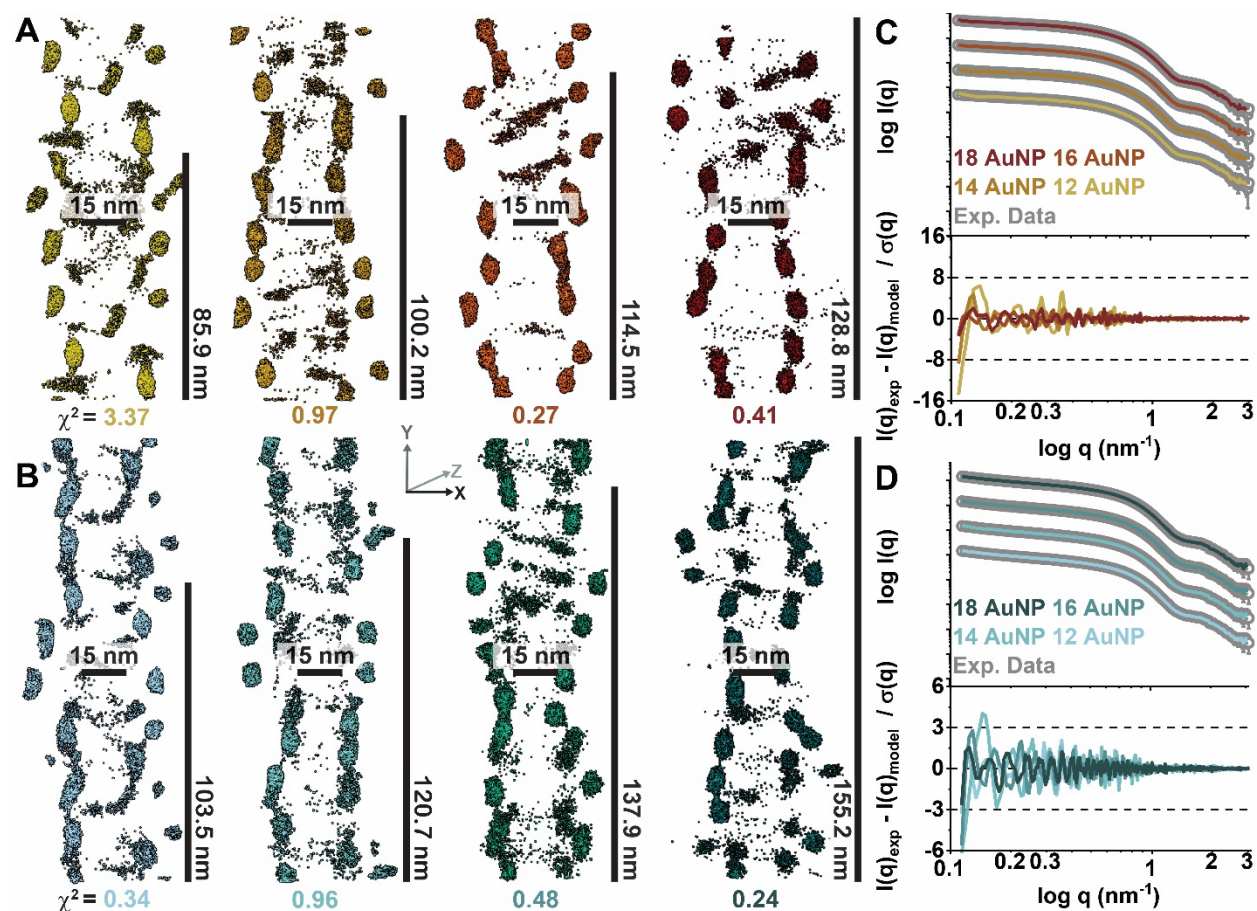
Figure 6-22. *Ab initio* modeling series results for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs made via SASHEL for 12 to 18 AuNPs per inter-AuNP block heights (H_{GB}), as defined in SI Methods Section 6.5 and Figure 6-19. χ^2-values are shown below each model and starting parameters are shown in Table 4-1. SAXS profiles with model fits and residuals for (C) (GT)$_{15}$-AuNP-SWCNTs and (D) (GT)$_6$-AuNP-SWCNTs, with each sample colored as in panel (A) and (B), respectively. Scattering curves are offset for clarity.
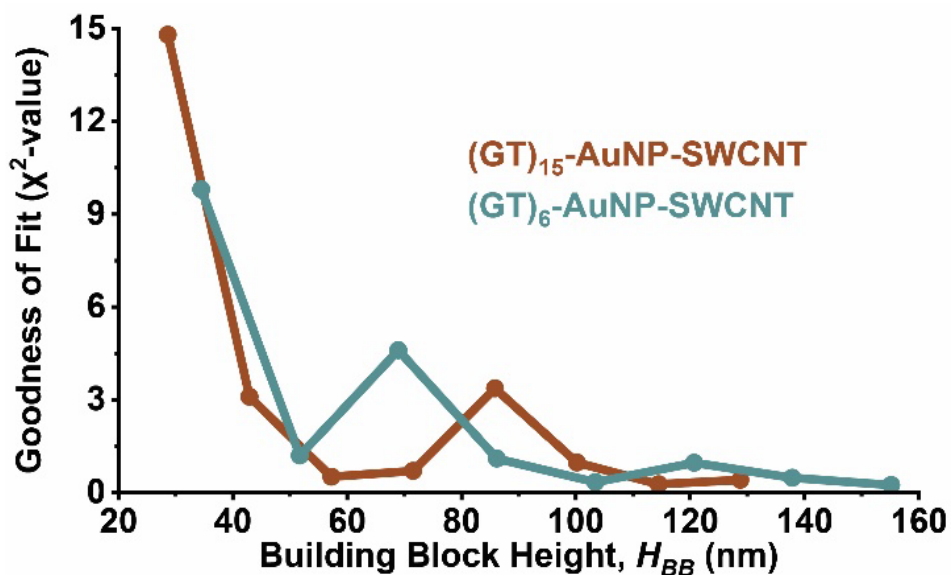
Figure 6-23. *Ab initio* modeling series results for (GT)$_{15}$-AuNP-SWCNTs and (GT)$_6$-AuNP-SWCNTs made via SASHEL showing the goodness of fit (χ^2-values) as a function of stack building block heights (H_{BB}). All parameters are shown in Table 4-1.
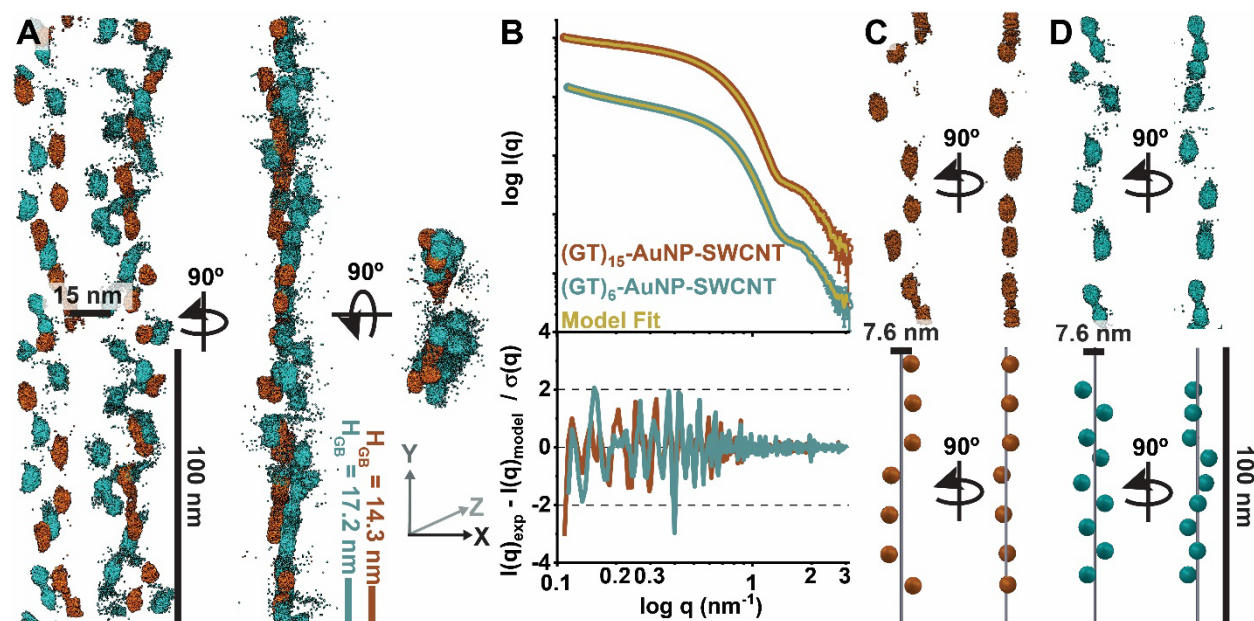
Figure 6-24. *Ab initio* modeling of ssDNA-AuNP-SWCNT complexes reveals periodic ordering of corona phase, using methodologies adapted from the SASHEL software for elongated nanoscale systems (see SI Methods Section 6.5). Both $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNTs were best captured by parallel lines of AuNPs (SWCNTs) with the (A) best-fit models for $(GT)_{15}$-AuNP-SWCNTs (red) and $(GT)_6$-AuNP-SWCNTs (blue) containing 16 or 36 AuNPs per repeating stack building block heights (H_{BB}), respectively. Additional parameters for inter-AuNP block heights (H_{GB}), number of stacks (N_S), and number of AuNPs (N_{NP}) used to produce the models can be found in Table 4-1. Noise reduction of the models was implemented for clarity by removing some un-clustered dummy atoms accounting for 8% of the total, as shown in Figure 4-3. (B) SAXS profiles with model fits (gold) and residuals for each complex colored as in panel (A). Scattering curves are offset for clarity. (C-D) Comparison of isolated, individual ssDNA-AuNP-SWCNTs from *ab initio* modeling (top) and theoretical 3D diagrams (bottom) produced from distances obtained from *P(r)* functions for (C) $(GT)_{15}$-AuNP-SWCNTs and (D) $(GT)_6$-AuNP-SWCNTs. Diagrams are scaled and colored to match that of panel (A).
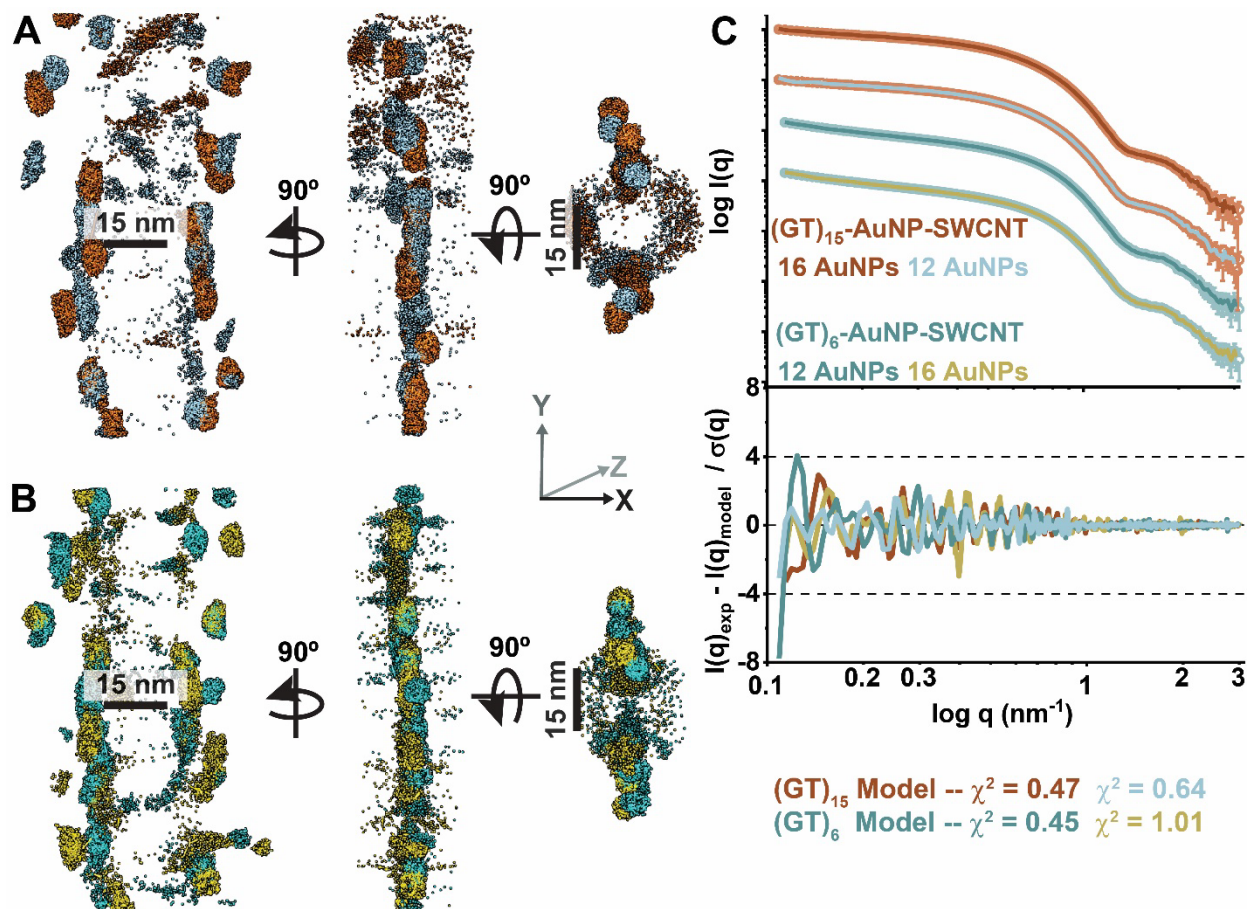
Figure 6-25. *Ab initio* modeling results for switching best-fit models for (A) (GT)$_{15}$-AuNP-SWCNTs and (B) (GT)$_6$-AuNP-SWCNTs. χ^2-values shown for each model. (C) SAXS profiles with model fits and residuals for (GT)$_{15}$- and (GT)$_6$-AuNP-SWCNT complexes with fits colored as in panel colored as in panels (A-B). Scattering curves are offset for clarity.
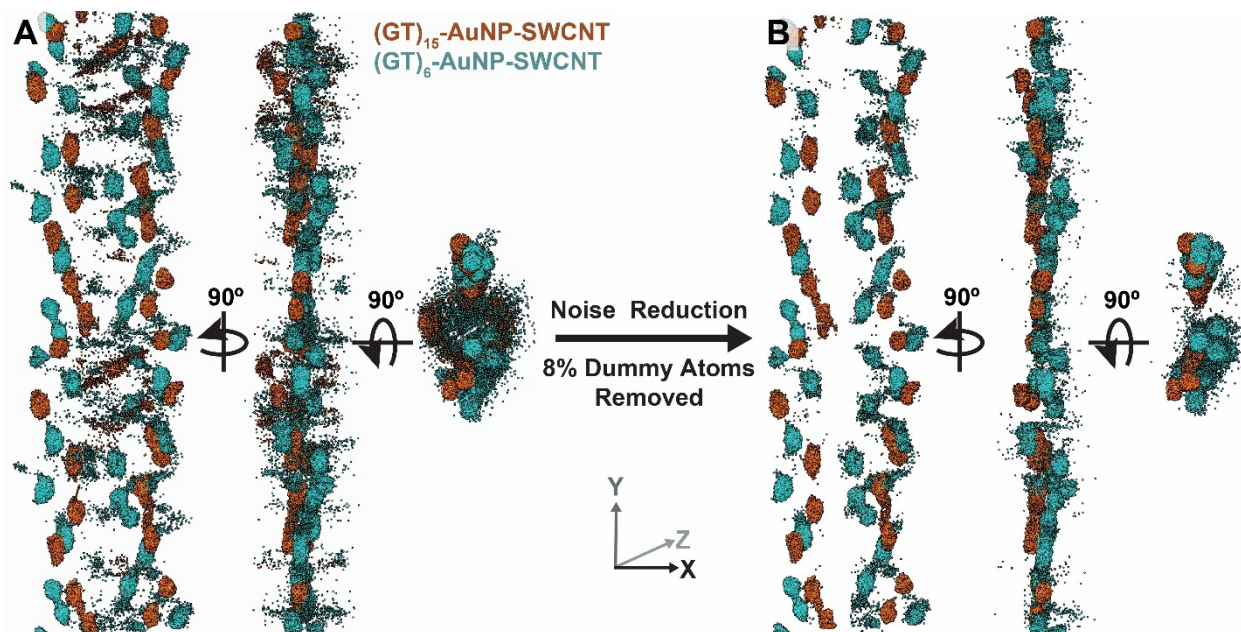
Figure 6-26. Noise reduction of best fit *Ab initio* modeling results by removing some un-clustered dummy atoms accounting for 8% of the total. Models are shown going from (A) original models for $(GT)_{15}$- and $(GT)_6$-AuNP-SWCNT to (B) noise reduced models.

7 Additional Cases for Size Exclusion Coupled Small Angle X-ray Scattering

7.1 Chapter Abstract

Since its inception in 2018, the size exclusion chromatography coupled SAXS with in-line multi-angle light scattering (SEC-SAXS-MALS) system has been responsible for more than 25 publications per year (see Chapter 2). This chapter highlights selected examples which the author feels best portray the capabilities of SEC-SAXS-MALS (most pertinent to this dissertation). Case 1: "Visualizing functional dynamicity in the DNA-dependent protein kinase holoenzyme DNA-PK complex by integrating SAXS with cryoEM", which highlights the ability of SEC-SAXS-MALS to complement higher resolution techniques (cryo-EM) while showing how it is powerful technique for studying DNA damage repair. Case 2: "Structural plasticity enables evolution and innovation of RuBisCO assembies", which demonstrates the robust high-throughput capabilities of the SEC-SAXS-MALS towards rational design strategies in bioengineering of molecular machines. Finally, case 3 and 4: "Rigid monoclonal antibodies improve detection of SARS-CoV-2 nucleocapsid protein" and "Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication" were specially selected as they were both projects made out of necessity at the beginning of the 2020 SARS-CoV-2 pandemic which at the time of filing this dissertation had resulted in 765,222,932 confirmed cases of COVID-19 and 6,921,614 deaths according to the most current data from the World Health Organization. Principally, these last two cases highlight the important role the SEC-SAXS-MALS technique can play in disaster responses on such magnitude.

7.2 Case 1: Visualizing functional dynamicity in the DNA-dependent protein kinase holoenzyme DNA-PK complex by integrating SAXS with cryoEM[††]

7.2.1 Introduction

A major enabling step in understanding non-homologous end joining (NHEJ) DNA repair has been realized through breakthrough comprehensive and high-resolution visualizations of DNA-PKcs by combining cryo-EM [264,265] with X-ray crystallographic [266] results. Defining the structural mechanisms for NHEJ is important for cancer etiology and therapeutic strategies since this is the primary DSB repair pathway in humans. Alternative DSB repair pathways either involve homology-directed repair or the more error prone alternative end joining that depends upon XRCC1 complexes [86,267-269]. To obtain structural insights before the current cryo-EM structures, it was necessary to build up models of the NHEJ complexes one component at a time [270]. Importantly, such insights pertain to genome instability seen in some cancers associated with translocations that depend upon the actions of NHEJ proteins and that are particularly associated with non-B DNA-forming sequences [271,272]. Moreover, as DNA-PK has functional interactions with poly(ADP-ribose) polymerase (PARP1) [273] and both PARP1 and poly(ADP-ribose)

---

glycohydrolase (PARG) inhibitors are actively being pursued for targeting the DNA damage response to kill cancer cells, these DNA-PK complexes are certainly relevant to ongoing cancer research [88,274].

Currently, in cases where cryo-EM provided only near atomic resolution, integration of high-resolution crystal structures of the assembly and components [266,275] into the cryo-EM maps enabled the reconstruction of atomistic models for the larger KU-DNA-DNA-PKcs (DNA-PK) assembly [276]. Yet, these are flexible complexes. In order to create a tractable sample for cryo-EM analysis, crosslinking agents were required to stabilize the complex. Such crosslinking limits assessment of flexibility but implies the complex is functionally dynamic. Indeed, the complex is expected to undergo allosteric transitions for function. We have therefore been developing and applying solution state small angle X-ray scattering (SAXS) techniques to complement cryo-EM and X-ray crystallographic structures, bridge size restrictions on techniques and provide a perspective on functionally relevant solution behavior.

Previously we employed SAXS data to evolve understanding of static structures into dynamic multistate functional conformations and to visualize flexible or unfolded regions [277-280]. Due to rotational averaging, the information content of SAXS is dramatically reduced compared to a density maps generated from cryo-EM or crystallography. Two advantages of SAXS, however, are that it provides objective data on flexibility [281] and that SAXS profiles can be efficiently calculated from atomistic models and directly matched to experimental data [282]. As a result, multistate data-based models that mimic dynamic rearrangements, such as domain motions, transient complexation, and the presence of unfolded regions, can be robustly determined by SAXS-based atomistic modeling [283,284]. Moreover, high-throughput SAXS data collection techniques, including size-exclusion, chromatography-coupled SAXS (SEC-SAXS)[9], enable the collection of many samples at multiple different conditions, facilitating the integration of SAXS with cryo-EM and crystallography [6,285,286]. E.g., such SEC-SAXS unveiled the basis for the regulation of the cancer-important P53 protein [287]. Indeed, first defining and then controlling functional conformations by small molecules can control activities and even repair pathway choice as seen for the MRE11 complex [73,268].

Here, we build upon published SAXS data and studies that characterized KU, KU-DNA, DNA-PKcs [286,288], and the DNA-PK assembly [288] and incorporate newly available atomic structures. Integrating atomic structures/models with new approaches in SAXS-based atomistic modeling [282,284,289] allowed us to derive dynamic models of the DNA-PK assembly and its components. From these analyses, we characterized a displacement of the KU80 C-Terminal Region (KU80CTR) from the KU core, show an inherent swing-like motion of the DNA-PKcs M and N-HEAT regions and identify a significant contraction of this N-HEAT region upon DNA-PKcs autophosphorylation. Interestingly, similar functional contractions and swing-like motions have been seen in the MRE11-RAD5O-NBS1 complex that acts in the homology-directed DSB repair as the major alternative pathway to NHEJ[269,290]. We moreover define and validate the solution state of the DNA-PK assembly that suggests stabilization of the DNA-PKcs HEAT regions upon recruitment to the KU-DNA complex. Intriguingly, we show that DNA-PKcs and DNA-PK can form a head-to-head interaction that appears to be a likely suitable precursor of the pre-synaptic NHEJ complex.

7.2.2   *Results*

7.2.2.1   KU80 C-terminal region (CTR) dynamicity in DNA-free and DNA-bound states

Enhanced SAXS experiments and analysis identify and verify a preferentially close interaction between the flexibly linked KU80CTR domain and the main KU70 α/β domain. Analysis of SAXS data collected on statics samples assuming homogeneous monomers in solution had suggested the complete disassociation of KU80CTR relative to the KU core in both DNA-free and DNA-bound states [286]. Our new measurements, obtained with size exclusion chromatography coupled with in-line SAXS and multi-angle light scattering (SEC-SAXS-MALS; see Chapter 2, Figure 7-1A, Figure 7-6), separated small amounts of transiently self-associating KU dimers, which had not been considered previously. Analyzing the SEC-SAXS from the monomeric peak demonstrated smaller radius of gyration (Rg) values and less elongated P(r) functions (Figure 7-1B, Figure 7-6, and Table 7-1) than those observed in the static experiment [286]. Further modelling through conformational sampling of the KU80CTR and KU80 C-terminal regions was applied [283] and, based on the goodness-of-fit (χ^2) between theoretical and experimental SAXS curves [291], the best single or multi-state models [284], were selected. The single state model showed a compact arrangement of KU80CTR neighboring the KU70 α/β domain (Figure 7-1C). Significant improvement in the SAXS fit was achieved by selecting the multistate model ($\chi^2_{\text{single state}}$ 4.2 vs. $\chi^2_{\text{two-state}}$ 2.2, Figure 7-1A). The multistate model included 55% of the conformers with the KU80CTR domain in close proximity to the KU70 α/β domain and 45% of the conformers with detached KU80CTR, ~30Å from the KU core (Figure 7-1C). The larger weighting of compact structures in solution validates and extends the models observed in cryo-EM studies showing relatively compact low resolution molecular envelopes, which suggested close contacts between KU80CTR and KU70 α/β domains [292] (Figure 7-1D). However, a significant population of detached KU80CTR may remain of mechanistic importance in recruiting of DNA-PKcs.

Next, SAXS analysis of the KU-DNA complex was prepared using 16-bp double-stranded DNA (dsDNA) with a short stem loop at one end as template [286]. Similar to prior SAXS analysis of the KU-DNA complex [286], only small differences between KU and KU-DNA measurements were observed. Particularly, narrowing of the P(r) functions (Figure 7-1B) indicates a less hollow structure that can be explained by insertion of DNA between the bridge and β barrel KU-regions [275]. Again, conformational sampling was performed on KU80CTR to determine the movement of KU80CTR relative to the KU-DNA core. In the search for representative multistate solution models, the KU-DNA experimental SAXS data was fit against a pool of ~10000 conformers including KU and KU-DNA models with various conformations of KU80CTR. Surprisingly, the best fit multistate model showed only 18% of the KU-DNA complex with the KU80CTR in close proximity to the KU core (Figure 7-1C). This partial occupancy of the KU-DNA complex suggests dissociation of KU from unprotected DNA. These results are in agreement with multiple assays under non-reducing condition [293].
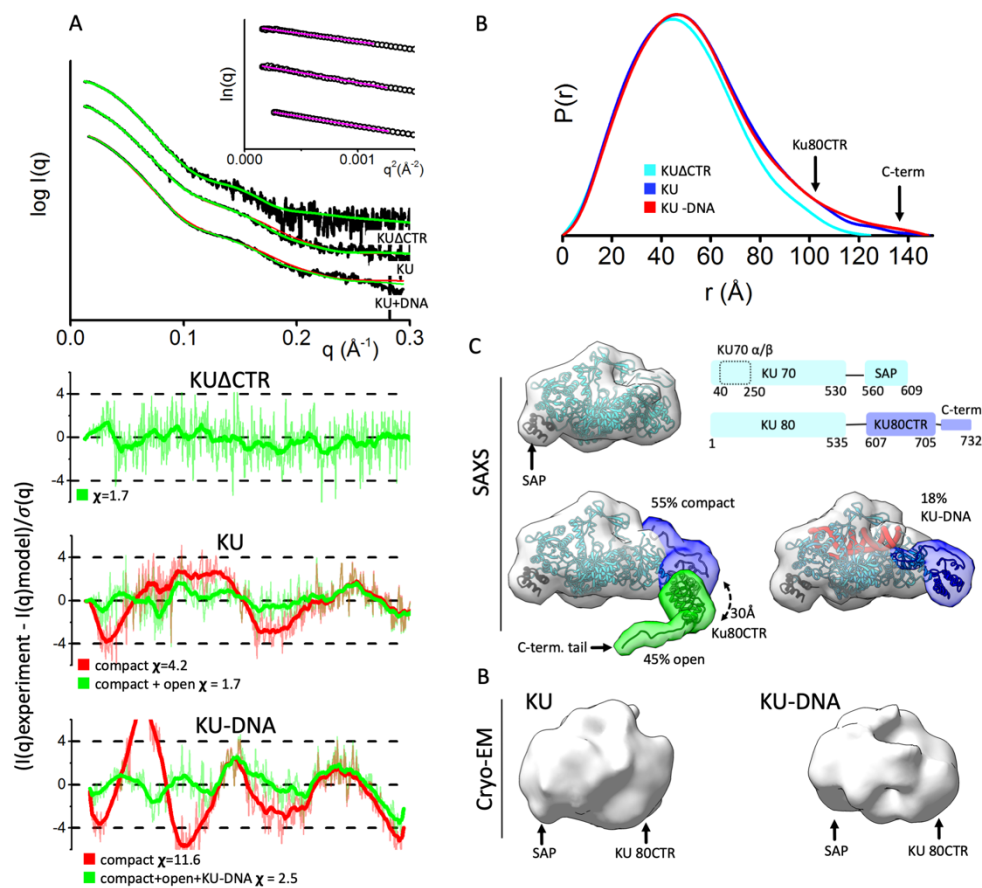
Figure 7-1. KU80CTR is located in close proximity to the KU70 α/β region in DNA free and DNA bound states. A) Experimental (black) and theoretical (colored as indicated) SAXS profiles for the solution state models of KUΔCTR, KU and KU-DNA. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in inset. B) Normalized pair distribution P(r) functions for experimental SAXS curves of KUΔCTR (cyan), KU (blue) and KU-DNA (red). C) top panel: Crystal structure of KU [275]. A schematic representation highlighting the domains of KU: two KU70 regions composed of the KU core region and the SAP domain; and three KU80 regions composed of the KU80 core region, the KU80CTR domain and the KU80 C-terminal helix. bottom panel: conformers in selected multistate model of KU and KU-DNA used to calculate theoretical SAXS. The weight of each model is indicated. The two conformers (compact and open) used to fit experimental SAXS curves of KU and KU-DNA. The KU-DNA conformer with 18% weight is shown independently. Atomistic models are displayed in ribbon style together with molecular envelop calculated at the 25Å resolution to match the resolution of cryo-EM maps shown at the bottom panel. D) cryo-EM maps of KU (EMD#:1270) and KU-DNA complex (EMD#1271) [292]

7.2.2.2 Conformational plasticity in the DNA-PKcs HEAT region

DNA-PKcs (4128 amino acids, ~469 kDa) [294] is composed of FAT, Kinase, FAT-C domains that form the "head" or crown and a large, flexible HEAT region, also called the palm region, that can be divided into the M-HEAT [295-297] and the recently visualized N-terminal HEAT region (N-HEAT)[264,266]. DNA-PKcs structures, together with previously reported cryo-EM low-resolution molecular envelopes [295,296] suggest that the DNA-PKcs HEAT region is flexibly attached to the "head" region allowing it to move during autophosphorylation [286] and rearrange upon interaction with the KU-DNA complex [276]. Our previous solution scattering studies have shown that DNA-PKcs undergoes self-association at higher protein concentrations [286]. Dimer-free DNA-PKcs SAXS data revealed a typical globular particle with a Dmax of 155 Å (Figure 7-2B and Table 7-1).

To test how the existing atomic resolution structures match the solution state, we compared existing DNA-PKcs structures to experimental SAXS curves. The available atomic models are missing loop regions corresponding to amino acids 2576-2776 which contains the ABCDE cluster of phosphorylation sites [298,299] located between the M-HEAT and FAT domain, and a conserved DNA-PK signature motif [300]. Without these domains, fits of the X-ray crystal [266] and Cryo-EM [264,276] structures were poor ($\chi2_{crystal}$ =128 crystal and $\chi2_{cryo-EM}$ =160) (Figure 7-2A). To further test whether this discrepancy was due to the flexibility of the missing regions, these sections were built using MODELER[301] and conformational sampling was applied using BILBOMD [283] to mimic the plasticity of the added regions. The goodness of SAXS fit improved by including flexibility in the ~2576-2776 region ($\chi2_{\text{single state}}$ =25.3); however, the remaining discrepancy between theoretical and experimental SAXS curves in the low-resolution range (q 0.05-0.15Å^{-1}, see Figure 7-2A), suggested larger conformational rearrangements of the DNA-PKcs domains in solution.

To visualize these allosteric changes, the conformational sampling protocol was applied using normal mode analysis (NMA) [289]. DNA-PKcs was divided into 4 regions (see Materials and methods) and constrained NMA sampling was applied to optimize the position of each region relative to each other including normal mode movement inside each region. The best SAXS fit model was found by searching the two-state models [284] from the pool of existing structures and 10 models derived from NMA conformational sampling (NMA-model). The best two-state model included 38% of the original structure and 62% of the NMA-model and was in excellent agreement with the experimental SAXS profile ($\chi2_{\text{two-state}}$ = 5.8, Figure 7-2A). The NMA-model showed a ~25Å displacement of the N-HEAT domain (Figure 7-2D) and agreed with the conformational variability of this region, as shown by comparison of the cryo-EM and crystal structures (Figure 7-2C). Additionally, the NMA-model showed a smaller ~10Å movement of the M-HEAT region. In sum, a multistate model of DNA-PKcs was determined experimentally that shows the flexibility of the HEAT region in solution.
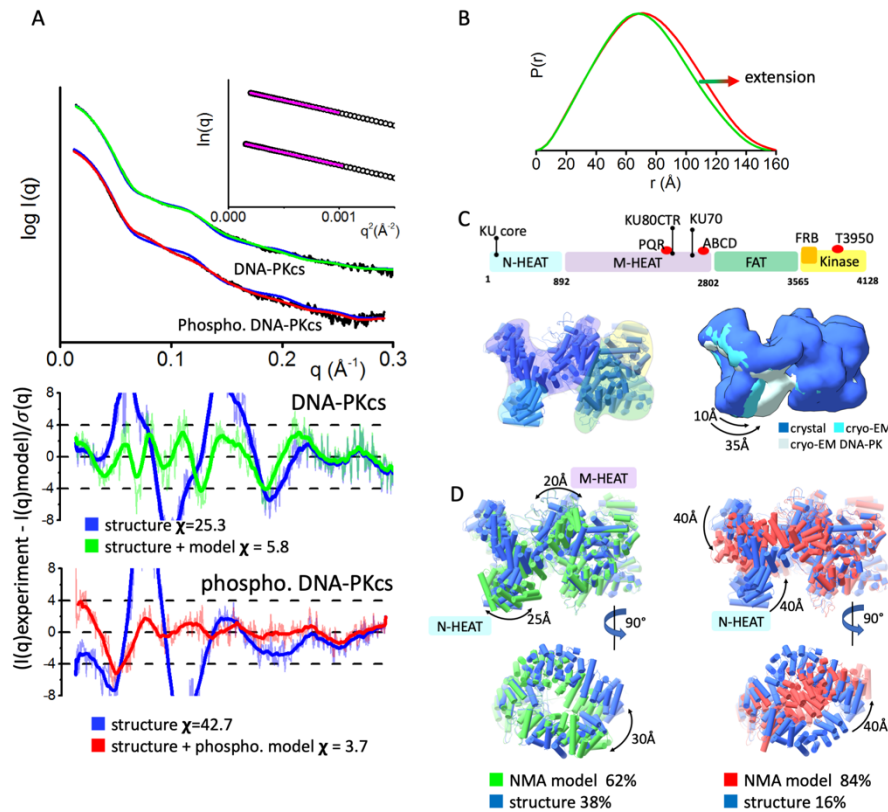
Figure 7-2. Inherent dynamicity of DNA-PKcs HEAT region and its rearrangement during the autophosphorylation. A) Experimental (black) and theoretical (colored as indicated) SAXS profiles for the solution state models of DNA-PKcs and autophosphorylated DNA-PKcs. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in the inset. B) Pair distribution P(r) functions, normalized at the maxima, for experimental SAXS curves of DNA-PKcs and autophosphorylated DNA-PKcs (taken from [286]). C) Top panel: A schematic representation highlighting the four super secondary structural components of DNA-PKcs: the two HEAT region composed of the N-terminal domain (N-HEAT); the M-HEAT region and the Head regions, which contains the FAT and kinase regions. The KU binding area, FRB domain, autophosphorylation clusters PQR and ABCD, and highly conserved T3950 autophosphorylation site are shown above the schematic. Left panel: Crystal structure of DNA-PKcs with highlighted N-HEAT, M-HEAT, FAT and kinase regions. middle panel: Comparison of the crystal structure and cryo-EM structure from [264,266], and cryo-EM structure of DNA-PKcs taken from the DNA-PK complex [276]. For better visualization of conformational variability in the HEAT region, atomic models are displayed as a molecular envelop at the 20Å resolution. D)Two orthogonal views of multi-state model used to match experimental SAXS curves of DNA-PKcs and autophosphorylated DNA-PKcs. The models were superimposed on each other at the FAT region. Weight for each model is indicated.

159

7.2.2.3 Displacement of the N-terminal HEAT domain during autophosphorylation suggests the release mechanism of DNA-PKcs from DNA-PK.

Autophosphorylation of DNA-PKcs prompts its release from KU-DNA double strand break (DSB) complexes *in vitro* and *in vivo* [286,302-305]. In prior studies, the *in vitro* autophosphorylated form of purified DNA-PKcs was isolated and analyzed using SAXS [286]. Comparison of the SAXS data of DNA-PKcs with and without autophosphorylation revealed different scattering profiles over the entire observed scattering range (Figure 7-2A). These changes were visualized in the P(r) function as broadening (Figure 7-2B) and indicated a large conformational change involving the relocation of DNA-PKcs domains rather than extension of a single domain or a local change at the phosphorylation site. Calculated SAXS profiles from atomic resolution DNA-PKcs structures disagree with experimental SAXS profiles ($\chi2_{cryo-EM}$ = 42.7). The large discrepancies seen between the theoretical and experimental profiles at low-resolution (q range 0.05-0.15Å-1, see Figure 7-2A), further suggest a large conformational rearrangement.

Using a similar process to our analysis of non-phosphorylated DNA-PKcs, constrained NMA conformational sampling was performed to modify the position of the N-HEAT, M-HEAT, FAT, and Kinase regions (see Materials and methods). The best fit model was found by searching the two-state models from the pool of existing structures and multiple NMA derived models. The best two-state model included 16% of conformations near the crystal structure [266] and 84% of the NMA-model. The two-state model fit was a great improvement over the single model fit and gave an excellent match to the experimental SAXS curve ($\chi2_{two-state}$ = 5.8, Figure 7-2A). The NMA-model showed large (~40Å) displacements of both the N- and M-HEAT regions leading to closure of the aperture between these domains (Figure 7-2D). Together with the extension of the M-HEAT region, the NMA-model explains the observed broadening of the P(r) function and the increasing Rg values relative to the non-phosphorylated DNA-PKcs. This atomistic model of autophosphorylated DNA-PKcs uncovers rearrangement of the entire HEAT region suggesting inaccessibility of the KU/N-HEAT binding site [276]. These results show that DNA-PKcs autophosphorylation conceals the interface between DNA-PKcs and KU. We hypothesize that by making the N-HEAT/KU binding site inaccessible, KU is forced to the detach from DNA-PKcs which subsequently allows other NHEJ processing enzymes, like DNA ligase Ligase IV and polynucleotide kinase phosphatase (PNKP), to bind the DSB.

7.2.2.4 Two DNA-PKcs bridged by 40bp DNA form a "dumbbell" arrangement.

Previously we showed similarity between the low resolution (33Å) DNA-PK cryo-EM map (EMD1210),[306] and SAXS envelopes of the dumbbell arrangement of two DNA-PKcs molecules bridged by a 40-bp duplex with a Y-shaped structure at one end (40bp Y-DNA) [286] (Figure 7-3). We also found that a more compact dumbbell dimer is formed in the presence of 40bp DNA with the hairpin (40bp H-DNA) and 40bp DNA with two blunt ends (40bpDNA) [286]. Distinct dumbbell arrangements in the presence of 40bpY-DNA were interpreted as a bridging of two DNA-PKcs with the bulky head regions pointing outward, whereas the more compact DNA-PKcs - 40bp H-DNA and DNA-PKcs - 40bp-DNA dimers were explained by a head-to-head arrangement.
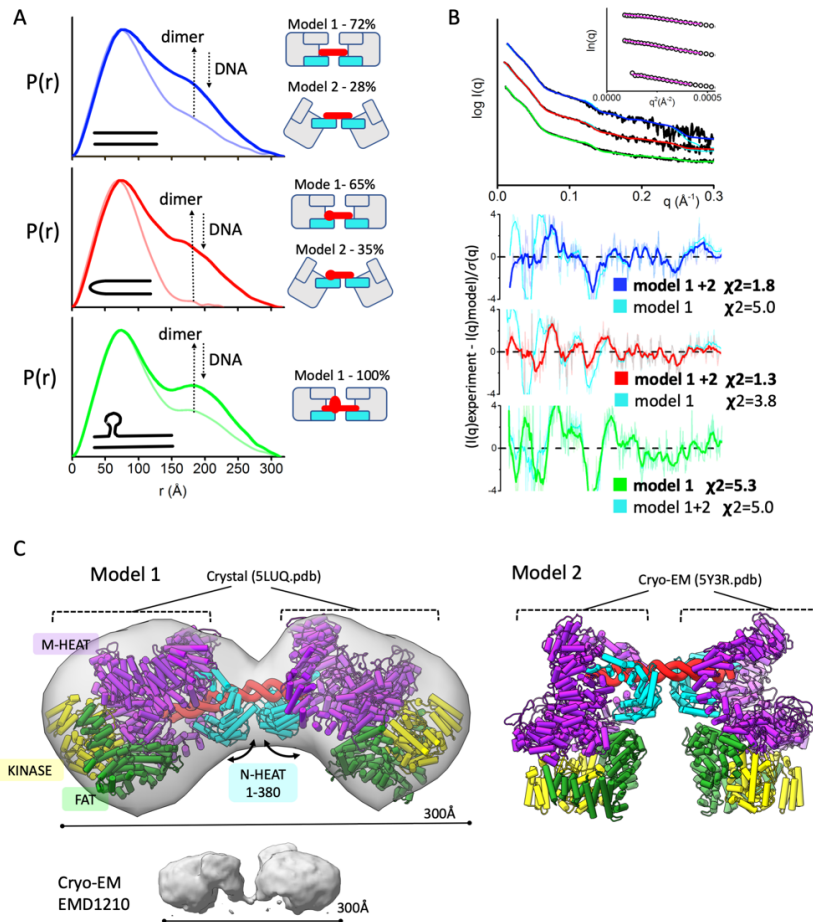
Figure 7-3. Formation of dumbbell DNA-PKcs-DNA dimers A) P(r) functions for DNA-PKcs - 40bp DNA (blue), 40bp H-DNA (red) and   40bp Y-DNA (green) with the ratio 2:1 (DNA-PKcs :DNA), calculated from the experimental SAXS shown in panel B. The light-colored P(r) functions are shown for the samples with equimolar DNA-PKcs:DNA molar ratio. The left panel shows cartoon representations of the atomistic models and its weights for each DNA-PKcs-DNA complex that were used to match the SAXS data shown in the panel B. B) Experimental (black) and theoretical SAXS profiles for the single (cyan) and multistate-model of DNA-PKcs in the complex with 40bp DNA (blue), 40bp H-DNA (red) and 40bp Y-DNA (green) with the DNA-PKcs:DNA ratio 2:1. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in the inset. C) Two atomistic models of DNA-PKcs dimer (model 1 and model 2) bridged by 40bpDNA(red). Model 2 was built based on the DNA-PK cryo-EM structure [276] by replacing KU with the DNA-PKcs. Model 1 was built   by replacing both DNA-PKcs with DNA-PKcs-crystal structure [266]. Conformational variability in the N-HEAT 1-380 region as seen between DNA-PKcs from DNA-PK (PDBID 5Y3R) and DNA-PKcs crystal structure (PDBID: 5LUQ) results in altering of DNA-PKcs tilt. The N-HEAT 1-380, M-HEAT, FAT and kinase regions are colored as indicated. cryo-EM map for putative DNA-PKcs-DNA-KU complex [306] is shown in the bottom panel.

Here we used SAXS with atomistic models to test formation of the dumbbell arrangement and consider the capability of DNA-PKcs to bind various DNA ends in the absence of KU. The DNA-binding site within the N-HEAT and M-HEAT aperture is rich in positively charged residues mostly located on the N-HEAT 1-380 region [276]. The lack of specific protein-DNA contacts and the opened DNA-binding site may allow DNA-PKcs to accommodate a wide spectrum of DNA ends [276]. Disappearance of the dimeric state in the excess of DNA (Figure 7-3A) [286] further supports the notion that the DNA-PKcs-DNA interaction is not stably maintained without KU [276,307]. By preserving the DNA binding site and mimicking conformational variability of the N-HEAT 1-380 region, DNA-PKcs can adopt different tilts relative to the linear DNA (Figure 7-3C). To fit the SAXS data, we used a pool of dimer models build based on DNA-PK structure [276] (see Methods Section 7.2.5). The best fit for the DNA-PKcs-40bp DNA dimer was obtained by a two-state model that includes 56% of the dimer using DNA-PKcs crystal structure (model 1) [266] and 44% of the dimer using DNA-PKcs taken from the DNA-PK structure (model 2) [276]. This two-state model gives a significant improvement in the SAXS fit over the single model ($\chi2$ one-state = 5.0 vs. $\chi2$ two-state = 1.8, Figure 7-3B). Model 1 and model 2 (65% and 35%) were also selected to give best SAXS fit of DNA-PKcs - 40bp H-DNA data with the significant improvement over the single model fit ($\chi2$ one-state = 3.8 vs. $\chi2$ two-state = 1.3, Figure 7-3B). On the other hand, the extended dimer of DNA-PKcs – 40bp Y-DNA was well matched with extended single model 1, whereas only a non-significant improvement in the SAXS fit was obtained by two-state model ($\chi2$ one-state = 5.0 vs. $\chi2$ two-state = 5.3, Figure 7-3B). These observations suggest a dominant presence of the extended dumbbell assembly: they further agree with well separated peaks in P(r) function (Figure 7-3A) and distinct volumes in the SAXS envelope (Figure 7-3C). Yet, this ability of DNA-PKcs to assemble on both ends of the DNA structure as visualized here and in previous studies [306] may not be physiologically relevant, as in vivo DNA damage-induced DSBs would each have only one exposed DNA end. In contrast, the self-association in the DNA-PKcs dimer provides an appropriate protein arrangement for the initial NHEJ step and is further described in the next section.

7.2.2.5   DNA-PK dimer reconstruction

In prior work [288], the overall DNA-PK complex architecture was characterized in solution with KU bound to a short dsDNA oligomer mimicking a DSB. For the current experiments, DNA-PKcs was mixed with KU that had been pre-incubated with 20 bp DNA duplex containing a short DNA stem-loop on one end and a 5'-nucleotide (nt) overhang on the other (20 bp DNA). SEC-SAXS-MALS measurements, complimented by SDS-PAGE analysis of the SEC fractions, confirmed the DNA-PK assembly. DNA-PK eluted as an asymmetric peak (Figure 7-3A), whereas MALS-analysis showed a decrease in the molecular weight (MW) of DNA-PK from 750 kDa at the beginning of elution peak to ~650 kDa at the tail (Figure 7-4A). Thus, under these experimental conditions, we found that DNA-PK was primarily forming a 1:1:1 assembly with a theoretical MW of 642 kDa. Analysis of SAXS frames across the primary elution peak also showed a decrease in Rg from ~75 Å at the peak to ~65 Å at the tail (Figure 7-4A, Table 7-1). These SEC-SAXS-MALS results suggest formation of a transient DNA-PK dimer that slowly interconverts over the elution time into a DNA-PK 1:1:1 assembly (DNA-PKcs/KU-DNA). The

formation of a DNA-PK dimer agrees with the tendency of DNA-PKcs to self-associate [286,308] (Figure 7-4C inset).

To further analyze both monomeric and dimeric solution states of DNA-PK, the primary SEC elution peak was deconvoluted into two SAXS profiles (peak and tail, Figure 7-4B). The P(r) function calculated for the SAXS curve at the elution peak showed an elongated assembly with a Dmax of ~300 Å, while the P(r) from the tail was narrower with a Dmax of ~220Å (Figure 7-4C). Due to the expected transient dimerization of DNA-PK, the SAXS curves collected at the peak and tail of the SEC were fit separately.

The monomeric DNA-PK model built based on a cryo-EM structure [276] matched well to the SAXS profile from the elution tail (Figure 7-4B). However, adding the missing parts of DNA-PKcs and KU, including the KU80CTR domain to the model improved the fit ($\chi2_{structure}$ = 8.0 vs. $\chi2_{full\ model}$ = 3.6) (Figure 7-4B). Further refinement of KU80CTR location relative to the KU core was not possible due to low signal-to-noise in the SAXS data. Nevertheless, the overall arrangement of DNA-PK supported and extended reported multiphase SAXS envelops [288] and further confirmed the overall DNA-PK architecture in solution (Figure 7-4D).

To build the DNA-PK dimer model, we initially modelled the DNA-PKcs dimer through the molecular docking of two DNA-PKcs monomers using a rigid docking, geometric shape-matching algorithm [309]. The best scoring model showed two DNA-PKcs monomers [266] in a mirror symmetry with a head-to-head arrangement (Figure 7-4E). Overall, the observed arrangement of self-associated DNA-PKcs resembled the previously reported V-like SAXS envelop [286] and low resolution 2D EM projection [308] (Figure 7-4E–inset). Furthermore, the docking model suggested contact between FKBP12-rapamycin-binding domains (FRB) which are conserved within the PI3K kinase family members (Figure 7-4E). While the role of the FRB domain is unclear, there is evidence from mTOR studies that it might act as a gatekeeper that restricts access to the catalytic site [310]. The stand-alone FRB domain appears to function as gate to the buried and conserved T3950 autophosphorylation site [266] that deactivates the kinase [302]. The dimerization interface of DNA-PKcs modeled in this study may alter FRB positioning and control access to the T3950 autophosphorylation site.

Building upon these results, atomic models of the DNA-PK dimer were constructed by aligning two DNA-PK monomers with the DNA-PKcs dimer. Exchange between the two states occurred over the course of purification, so that a pure dimeric state could not be isolated from the monomeric one. Thus, the SAXS curve generated from the SEC peak was fit against a library of atomic models including DNA-PKcs structures, DNA-PKcs models, DNA-PK monomer, and DNA-PK dimer. The best fit two-state model was obtained by including 24% of DNA-PK dimer and 76% of DNA-PK monomer conferring an excellent match to the experimental SAXS profile ($\chi2$ = 0.9, Figure 7-4BE). The SAXS modeling together with the determined MW demonstrated the tendency of DNA-PK to form a head-to-head dimer with a V-shape arrangement. In the model, KU-DNA was positioned at the extremity of the dimer with the DNA entering an aperture in DNA-PKcs between the N- and M-HEAT regions.
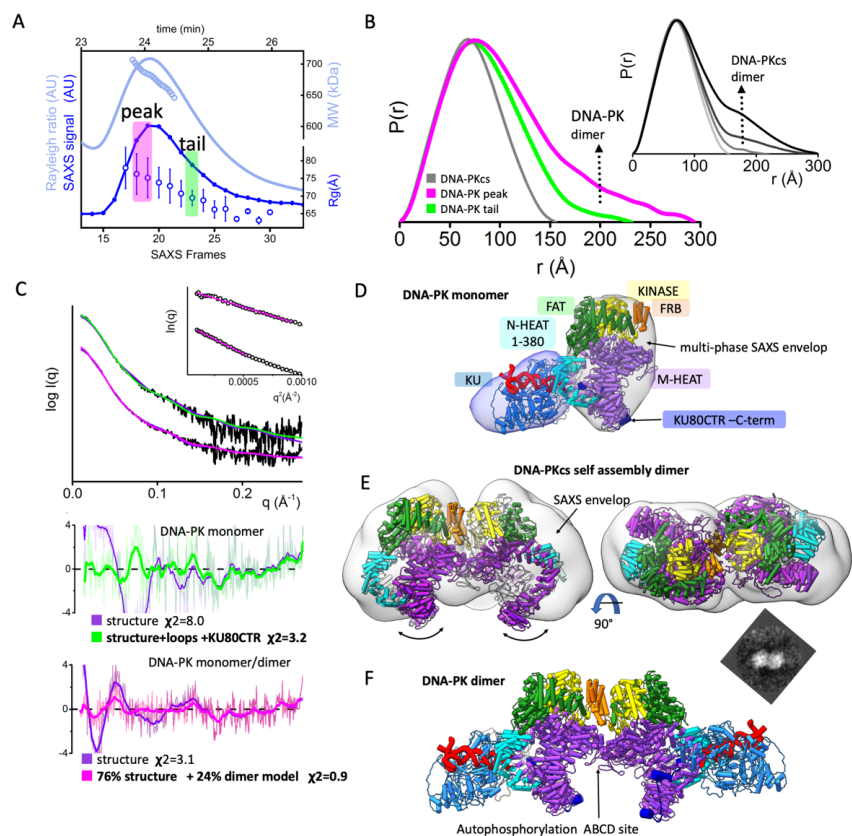
Figure 7-4. Solution state of DNA-PK and its dimerization through FRB domain. A) SEC-MALS-SAXS chromatograms for DNA-PK assembly. Solid lines represent the MALS signal shown as Rayleigh signal (light blue) or integrated SAXS signal (dark blue) in arbitrary units, while symbols represent molecular mass (light blue) and Rg values for each collected SAXS frame (dark blue) versus elution time (taken from [288]. B) Experimental (black) and theoretical (colored as indicated) SAXS profiles for the solution state of DNA-PK in monomeric and dimeric state. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in inset. C) Normalized pair distribution P(r) functions for experimental SAXS curves of DNA-PK assemblies measured at the peak and tail (green) of the elution peak (magenta) in comparison to monomeric DNA-PKcs taken from [286]. Inset: Normalized P(r) functions calculate for the experimental SAXS curves of DNA-PKcs collected at protein concentrations 1.5, 3, 7 and 15 mg/ml (from light gray to black) indicates self-association of DNA-PKcs at higher concentrations (taken from [286]). D) The cryo-EM structure of DNA-PK [276] is superimposed on to the multi-phase SAXS envelop of DNA-PK taken from [288]. E) Two orthogonal views of DNA-PKcs docking model representing the self-association dimer. The DNA-PKcs self-association dimer superimposed on to the SAXS envelop reconstruction for the SAXS data collected at the 15 mg/ml protein concentration (taken from [286]. Additional top scoring models are shown in the Figure 7-7. The 2D EM projection of DNA-PKcs self-association dimer (taken from [308]). F) Proposed atomic model of DNA-PK – dimer that was used in 24% weight to match the SAXS curve measured at the SEC elution peak (see panel A).

7.2.3   Discussion

The promise of macromolecular structural biology is to visualize structures that unveil critical functional mechanisms. This promise is too often not fully realized due to missing or incomplete knowledge of functional conformations assessed outside of cryo- and solid-state conditions [278]. On the other hand, SAXS, which can be high-throughput and directly measures thermodynamic solution-state conformational states and assemblies, is resolution limited. Yet, SAXS can be combined with other measurements to interrogate atomic level information [311]. Here we assess practical and robust methods to join SAXS with cryo-EM and X-ray crystallographic data as applied to the dynamic structure of DNA-PKcs and its interactions with KU and DNA in solution. The results reported provide insights into the dynamic architectural changes whereby these complexes orchestrate non-sequential NHEJ repair [312-314]. Previously our SAXS results predicted XRCC4 and XLF protein filaments for DNA end protection and alignment in DSBR[315,316], and these filaments were subsequently identified in single-molecule experiments as well as in cells [317,318]. Importantly, the new results presented here uncover displacement of the N-terminal HEAT domain during autophosphorylation as suitable for a regulated release mechanism of DNA-PKcs from DNA-PK to control unproductive access to toxic and mutagenic DNA repair intermediates. Such release of unproductive complexes has been seen before for FEN1 and XPG nuclease functions in base and nucleotide excision repair [50,319,320]. Overall, we present new SAXS data, revisit solution structure modeling of the KU heterodimer, and combine SAXS measurements with atomic resolution static structures.

 Our improved experiments on KU reveal that a major population of KU80CTR is mostly located in close proximity to the KU core. However, a nearly equal part (45%) of the solution state involves an extension of KU80CTR domain through the flexible linker (Figure 7-1C) that is in apparent disagreement with the static picture of KU visualized by cryo-EM [292]. Yet, our results support and extend earlier studies showing proteolytic sensitivity of the KU80CTR [321,322] and subsequent structural studies revealing a disordered linker region [323]. We reason that the detachment of KU80CTR from KU core remains functionally important for its role in the recruiting of DNA-PKcs. By superimposing DNA-PK and DNA-PKcs-KU80CTR structures [266,276], we can determine the distance between the KU core and KU80CTR (Figure 7-4D). When KU is bound to DNA-PKcs, the KU80CTR region needs to be far more extended from the KU core (~60Å) than in the free state (Figure 7-5). Additionally, the C-terminal interaction motif of KU80CTR [324], which is predicted to have helical propensity, is even more distant (~80Å) from the KU core [266]. Thus, the KU80CTR domain including the KU80CTR C-terminus needs to undergo a large displacement during KU interaction with DNA-PKcs. Such a dramatic rearrangement is enabled by the length of the KU80CTR linker (~60 residues). To initiate complexation, the flexibly tethered KU80CTR C-terminus helix needs to find the binding site near the "PQR" autophosphorylation cluster [266,325]. This initial KU-tethering is followed by recruitment of the KU core to the N-HEAT binding site, allowing insertion of the DNA end into the M- / N-HEAT aperture (Figure 7-5). The relatively compact arrangement of KU80CTR in the absence and presence of DNA suggests that this KU80CTR rearrangement is initiated by interaction between the KU80CTR C-terminus and the M-HEAT domain [266] rather than by DNA binding (Figure 7-5).

165

Alone, DNA-PKcs interacts with DNA with low affinities [307]. However, its ability to assemble on various ends of the DNA structure can lead to formation of dumbbell arrangement with two DNA-PKcs binding DNA ends [286]. This DNA-PKcs dimer arrangement is evidently not physiologically relevant, because in vivo DNA damage induced DSBs would each have only one exposed DNA end. A tighter interaction between DNA-PKcs and DNA is supported by KU (Figure 7-5) [276,286,305]. DNA-PKcs and KU together form a DNA-binding tunnel where DNA duplex fills the aperture between the N-HEAT and the M-HEAT region of DNA-PKcs [276]. We find that in solution the N-HEAT region undergoes large motions in free DNA-PKcs (Figure 7-5), in agreement with the conformational variability between reported DNA-PKcs structures [264-266] (Figure 7-2C), as well between two DNA-PKcs chains from the crystal structure [266]. Larger movement of N-HEAT and M-HEAT regions in solution (Figure 7-2D) correlate with much smaller changes in conformation and position of the HEAT regions observed in the crystal and cryo-EM structure. In the DNA-PK structure [276] the N-HEAT domain is displaced by ~35Å inwards towards the FAT domain upon binding with KU-DNA (Figure 7-2C). Our solution-based modeling shows similar movement of the N-HEAT domain in the absence of KU-DNA (Figure 7-2D), which supports the importance of inherent flexibility in this domain upon DNA-PK activation. Conformational adaptability of the N-HEAT region, bending of KU70 [276], and bridging of the KU80CTR C-terminus with the M-HEAT domain all contribute to a lock-in mechanism that stabilizes the formation of the DNA-PK assembly (Figure 7-5). Such a mechanism would ensure that assembly of DNA-PK controls a molecular switch needed to transfer the activation signal from the KU80CTR binding site to the kinase located in the head region ~100Å away (Figure 7-5).

Formation of a small but significant population of dimeric DNA-PK in solution (Figure 7-4B) suggests that self-association of DNA-PKcs [286] (Figure 7-4E) may support the dimerization of the DNA-PK complex. Here modeled head-to-head arrangement of the DNA-PKcs dimer resembles the arrangement of the closely related ATM kinase dimer [326]. This arrangement raises the question of whether the FRB dimerization controls the autophosphorylation, which activates the release of DNA-PKcs from the NHEJ presynaptic complex [299,303,327].

The main rearrangement of DNA-PKcs during autophosphorylation shows widening of the M-HEAT region (Figure 7-2D) that allows the N-terminal domain (1-370) of the N-HEAT to slide in to the cradle formed by wider M-HEAT region (Figure 7-5). We hypothesize that this mechanism may conceal the binding site for KU70 (Figure 7-5) and trigger release of DNA-PKcs from the DNA ends.

Autophosphorylation-induced domain rearrangement furthermore has the potential to affect the interaction of DNA-PKcs with accessory proteins, which may further regulate NHEJ *in vivo*. These collective results and ideas support and extend our previous structural understanding of DNA-PK activation and flexibility. Specifically, integrating SAXS measurements with the atomistic modeling utilizing crystal and cryo-EM structures provides data-based dynamic models that suggest how KU and DNA-PKcs combine and assemble enzymatically to promote structural and catalytic activities for NHEJ initiation and the choreography of DSB repair. Going forward will be interesting to see how RNA may regulate or active these DNA-PK complexes [328]. Furthermore, in this and other efforts to combine advanced cryo-EM with SAXS, the use of gold-label SAXS may provide a means to reduce SAXS samples needs substantially [34].
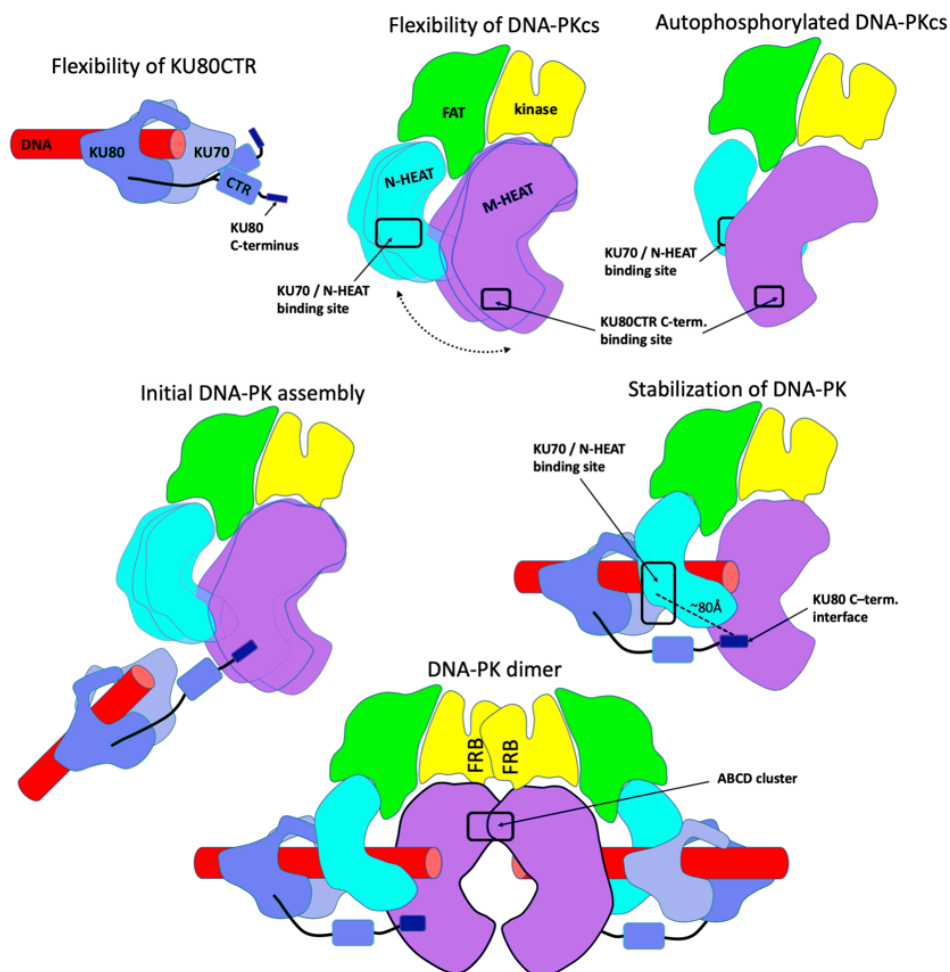
Figure 7-5. DNA-PK functional flexibility as cartoons. i) Minimal flexibility of KU80CTR and extension of its C-terminus. ii) Flexibility of HEAT domain relative to the DNA-PKcs head region. iii) N and M-HEAT rearrangement upon the DNA-PKcs autophosphorylation. iv) Extension of the KU80 "arm" during initial DNA-PK assembly. v) Stabilization of the DNA-PK assembly by multiple protein-protein and protein-DNA interactions. vi) Proposed DNA-PK dimer arrangement.

7.2.4   Supplementary Material

Table 7-1. Structural parameters from SAXS and MALS data.

SAXS sample SASBDB#	D_{max} (Å)	Rg (Å) from Guinier plot	Rg (Å) from P(r)	MW Seq. monomer (kDA)	MW SAXS (kDa)	MW MALS (kDa)	Model fit χ2	data source
KUΔCTR SASDJV4	~ 125	37.9±0.6	37.9	134	~ 140	145	1.7	SEC-SAXS (This study)
KU SASDJU4	~ 155	41.5±1.4	42.7	153	~ 160	160	1.7	SEC-SAXS (This study) merged with SAXS from [286]
KU-DNA SASDJW4	~ 155	40.7± 0.3	41.9	173	~ 170	190	2.6	Data from [286]
DNA-PKcs SASDJX4	~ 155	57.1±1.7	54.4	469	~ 480	480	5.9	SEC-SAXS from [30] merged with SAXS from [286]
Autophospho. DNA-PKcs SASDJY4	~ 160	57.2±1.7	56	469	~ 480	ND	3.7	SAXS from [286]
DNA-PK monomer SASDJZ4	~ 230	65.1±0.6	66.8	640	~ 590	640	3.7	SEC-SAXS from [288]
DNA-PK monomer/dimer SASDJ25	~ 300	75.4±1.5	83.5	640	~ 720	680-700	1	SEC-SAXS from [288]
DNA-PKcs – 40bp DNA	~ 315	80.9±4.2	92.9	499	~860	ND	1.8	SAXS from [286]
DNA-PKcs – 40bp H-DNA	~ 315	87.7±6.3	96.4	497	~900	ND	1.3	SAXS from [286]
DNA-PKcs – 40bp Y-DNA	~ 315	90.2±4.1	101.2	501	~880	ND	5	SAXS from [286]
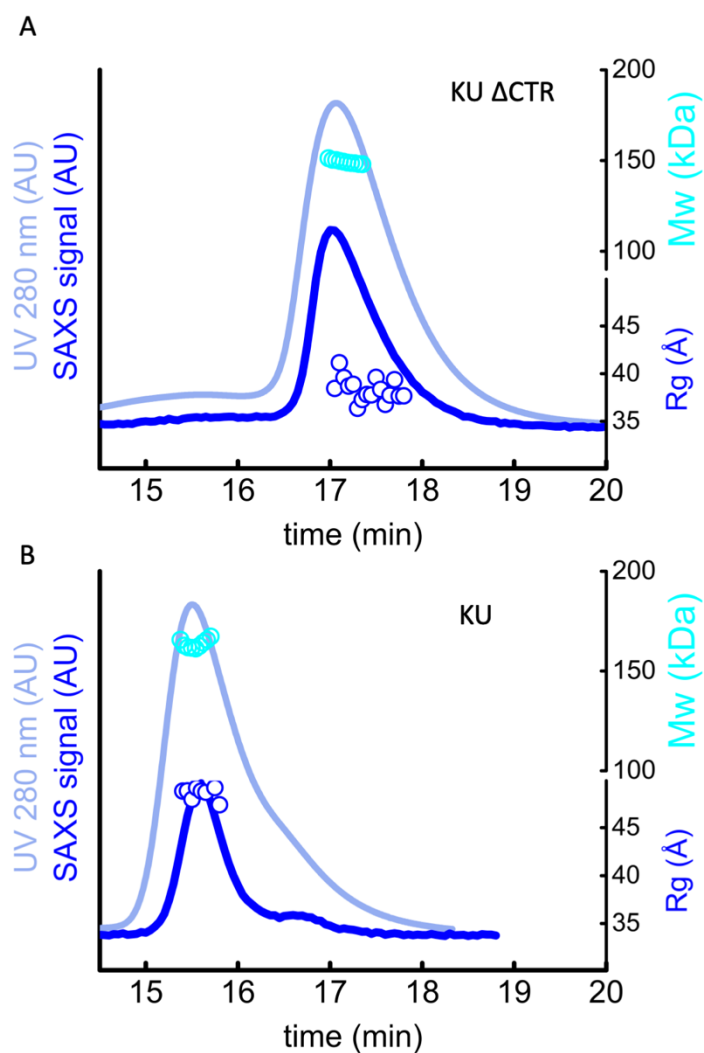
Figure 7-6. A-B) SEC-MALS-SAXS chromatograms for KUΔCTR and KU assembly. Solid lines represent the MALS signal shown as UV signal (light blue) or integrated SAXS signal (dark blue) in arbitrary units, while symbols represent molecular mass (light blue) and Rg values for each collected SAXS frame (dark blue) versus elution time.
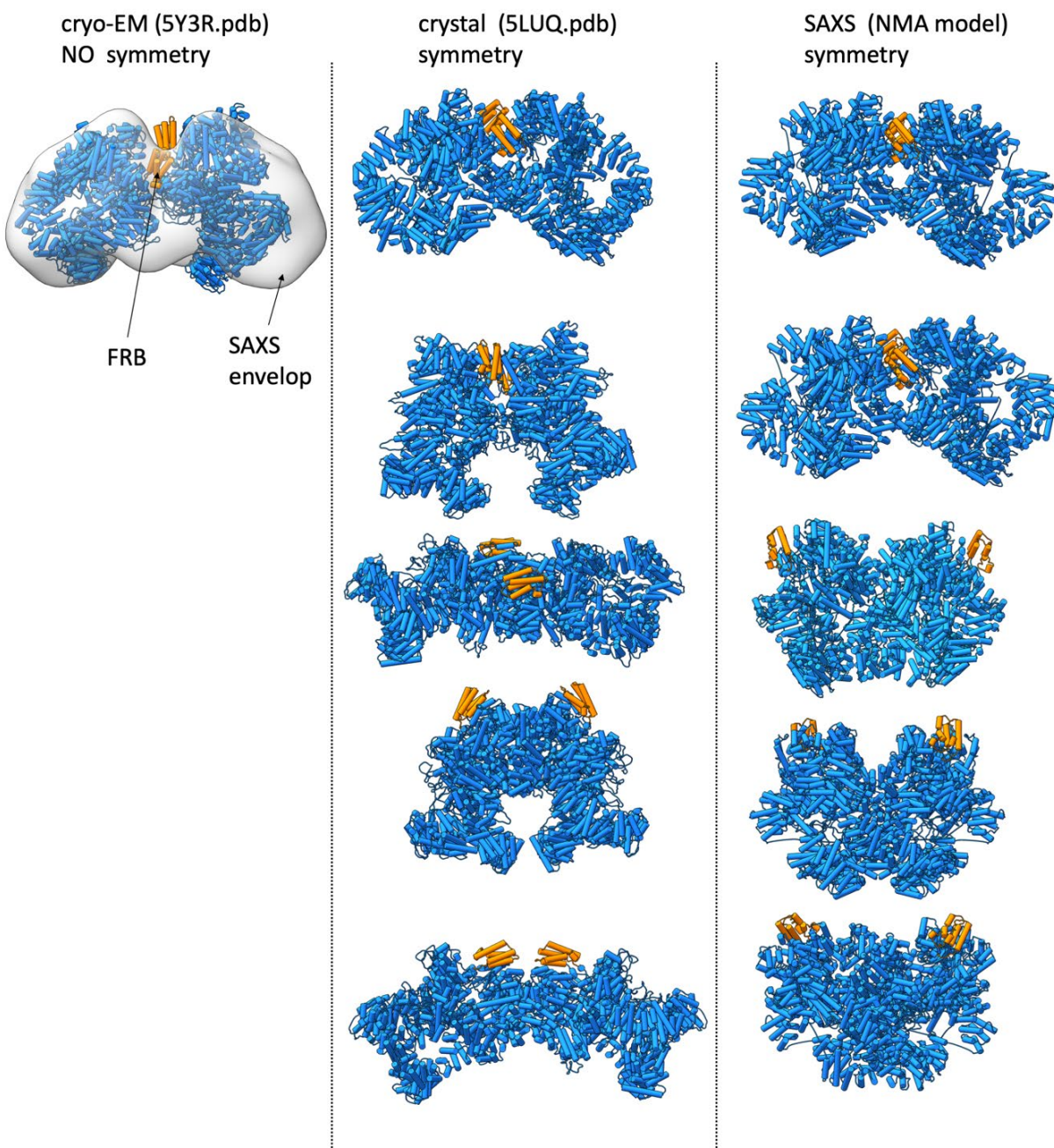
Figure 7-7. Left panel - The dimer of DNA-PKcs reconstructed by docking of two DNA-PKcs taken from the DNA-PK structure (PDBID: 5Y3R[276]) without symmetry operator. SAXS envelop for self-association DNA-PKcs dimer, taken from [286] is superimposed on the dimer model. Middle panel - The dimer of DNA-PKcs reconstructed by docking of two DNA-PKcs crystal structures [266] (PDBID: 5ULQ) using symmetry operator. Right panel - The dimer of DNA-PKcs reconstructed by docking of two SAXS-based DNA-PKcs models using symmetry operator. FRB domains are colored orange.

7.2.5   *Materials and Methods*

7.2.5.1   Purification of proteins

DNA-PKcs and KU were purified from HeLa cells as described [329,330]. KUΔCTR (KU70/KU80 1-569) were expressed and purified from baculovirus-infected insect cells as described [331]. Large-scale preparation of phosphorylated DNA-PKcs for structural analysis were prepared as described by [286]. 20-bp double-stranded DNA containing a short DNA stem-loop on one end for the preparation of KU-DNA complex was annealed and purified according [286]. 20-bp double-stranded DNA containing a short DNA stem-loop on one end and a 5'-nucleotide (nt) overhang on the other for the preparation of the DNA-PK complex was annealed and purified according[288].

7.2.5.2   SAXS experiment

For small-angle X-ray scattering coupled with multi-angle light scattering in line with size-exclusion chromatography (SEC-SAXS-MALS; see Chapter 2) experiments, 60 µL samples containing either 5 mg/mL of KUΔCTR and 5 mg/mL KU were prepared in 50 mM Hepes 7.5, 50 mM KCl, 5 mM $MgCl_2$, 5% glycerol and 0.2 mM DTT. SEC-SAXS-MALS were collected at the ALS beamline 12.3.1 LBNL Berkeley, California [332]. X-ray wavelength was set at λ=1.127 Å and the sample-to-detector distance was 2100 mm resulting in scattering vectors, q, ranging from 0.01 Å-1 to 0.4 Å-1. The scattering vector is defined as q = 4πsinθ/λ, where 2θ is the scattering angle. All experiments were performed at 20°C [333] and data was processed as described [9]. Briefly, a SAXS flow cell (see Section 2.3.2) was directly coupled with an online Agilent 1260 Infinity HPLC system using a Shodex KW803 column. The column was equilibrated with running buffer with a flow rate of 0.5 mL/min. 55 µL of each sample was run through the SEC and 3 second X-ray exposures were collected continuously during a 30min elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by the radius of gyration R_g derived by the Guinier approximation $I(q) = I(0) \exp(-qRg)^2/3$ with the limits $qR_g<1.5$. The elution peak was mapped by comparing integral of ratios to background and R_g relative to the recorded frame using the program SCÅTTER (Figure 7-6). Eluent was subsequently split 3 to 1 between SAXS line and a series of UV @ 280 and 260 nm, multi-angle light scattering (MALS), quasi-elastic light scattering (QELS), and refractometer detector. MALS experiments were performed using an 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab refractive index concentration detector (Wyatt Technology). System normalization and calibration was performed with bovine serum albumin using a 45 µL sample at 10 mg/mL in the same SEC running buffer and a dn/dc value of 0.19. The light scattering experiments were used to perform analytical scale chromatographic separations for MW determination of the principal peaks in the SEC analysis. UV, MALS, and differential refractive index data was analyzed using Wyatt Astra 7 software to monitor the homogeneity of the sample across the elution peak complimentary to the above-mentioned SEC-SAXS signal validation (see Figure 7-6).

The DNA-PKcs and KU-DNA-DNA-PKcs (DNA-PK) complex was prepared and measured by SEC-SAXS-MALS experiment (see Figure 7-3A) as described by [288]. Additionally, SAXS data of KU, KU-DNA, DNA-PKcs, phosphorylated DNA-PKcs were measured by high-throughput

SAXS experiment (HT-SAXS) as described by [286]. To improve the signal to noise ratio at higher q range SAXS curved derived from SEC-SAXS and HT-SAXS experiments were merged for KU and DNA-PKcs samples (see Figure 7-1A and 2A, Table 7-1). DNA-PKcs in complex with 40-bp duplex with a Y-shaped structure at one end (DNA-PKcs-40bp Y-DNA) and 40bp DNA with the hairpin (40bp H-DNA) [286] and 40bp DNA with two blunt ends were measured by high-throughput SAXS experiment (HT-SAXS) as described by [286].

7.2.5.3   SAXS data evaluation.

Final merged SAXS profiles where used for further analysis including Guinier plot which determined an aggregation free state (see Figure 7-1A, Figure 7-2A, and Figure 7-3B). The program SCÅTTER was used to compute the pair distribution function P(r). The distance r where P(r) approaches zero intensity identifies the maximal dimension of the macromolecule (D_{max}). P(r) functions of KUΔCTR, KU, KU-DNA (Figure 7-1B) and DNA-PK (Figure 7-3C) were normalized based on the molecular weight of the assemblies as determined from SAXS curves by SCÅTTER using volume of correlation V_c (see Table 7-1) [334,335]. To better visualized broadening of phosphorylated DNA-PKcs, the P(r) functions for DNA-PKcs and phosphorylated DNA-PKcs were normalized at their maxima (see Figure 7-2B).

7.2.5.4   Solution Structure Modeling

The pool of KU conformers from our previous study [286] was used to fit the experimental SAXS curve of KU using FOXS [284,291,336] and a multistate state model was selected using MultiFOXS [284] (see Figure 7-1C). The same pool of conformers plus the pool of KU-DNA models from [286] was used to fit experimental SAXS of KU-DNA. A three-state model was selected by MultiFOXS [284] (see Figure 7-1C).

To fit SAXS of DNA-PKcs and phosphorylated DNA-PKcs we initially built missing loops, including the ~2576-2776 region that contains the ABCDE phosphorylation sites [298,299] in to the X-ray crystal [266] and Cryo-EM structures[264] using MODELLER [301]. First we applied conformational sampling BILBOMD [283] to mimic the plasticity of the added loops. Next we optimized SAXS fit by large movement of DNA-PKcs domains using normal mode analysis (NMA) program SREFLEX [289]. The SREFLEX program uses NMA in Cartesian space to estimate the flexibility of atomistic models to improve their agreement with experimental SAXS curve. The DNA-PKcs crystal structure [266] with added missing loops was divided into 4 regions that include residues 1-370 (N-terminal N-HEAT region), 371-1800 (N-HEAT and M-HEAT), 1801-2800 (second region of M-HEAT) and 2801-4119 (FAT and Kinase region). SREFLEX optimized the position of each domain relative to the others including normal mode elastic movement in the region.   The generated 8 NMA models plus the crystal [266] and Cryo-EM [264] structures (with added missing loops) is a pool of the models that were fitted to the experimental SAXS by FoXS [284,291,336]   followed by selection of two-state model by MultiFOXS [284].

To fit the SAXS data of DNA-PKcs- 40-bp duplex with a Y-shaped structure at one end (40bp Y-DNA), 40bp DNA with the hairpin (40bp H-DNA) and 40bp DNA with two blunt ends (40bpDNA) we built atomistic model of DNA-PKcs dumbbell dimer by replacing KU in the

DNA-PK structure [276] with DNA-PKcs (model 2). Alternative model with the different conformation of N-HEAT 1-380 region were built by replacing DNA-PKcs with the conformers as seen in the crystal structure (model 1) (Figure 7-3C). Due to conformational flexibility of N-HEAT 1-380 region the DNA-PKcs may adopts different tilt relative to the linear DNA (Figure 7-3C). Thus, alternative models with various tilts of DNA-Pcs were built (Model 3-9). We used the pool of the models that include dimer-models and DNA-PKcs monomer to fit the SAXS curve for all three DNA complexes using program FoXS [284,291,336], followed by selection of two-state model by MultiFOXS [284].

To fit the SAXS curve of the monomeric DNA-PK assembly we initially built missing loops in the Cryo-EM structure [276] using MODELLER [301]. The missing KU80CTR domain was added in the close proximity of previously identify KU80CTR binding site at the M-HEAT region [266]. DNA-PK model was fitted to the experimental SAXS data by FoXS [284,291,336].

To fit experimental SAXS curves of the DNA-PK monomer/ dimer mixture we initially modeled a DNA-PKcs dimer by molecular docking of two DNA-PKcs monomers using a rigid docking, geometric shape-matching algorithm PatchDock [309]. The docking of DNA-PKcs (taken from DNA-PK structure PDBID: 5Y3R [276]) was performed without symmetry operator. The docking of the DNA-PKcs crystal structure [266] (PDBID: 5ULQ) and SAXS-based atomistic model was performed with symmetry operator. The top scoring model derived without symmetry operator resemble top scoring models with symmetry operator and was further used to build the DNA-PK dimer model (Figure 7-7). DNA-PK dimers were constructed by aligning two DNA-PK monomers with the top scoring docking model of DNA-PKcs dimer. Experimental SAXS curve of DNA-PK monomer/ dimer mixture was fitted by both monomer and dimer model using FoXS [284,291,336] followed by selection of two-state model by MultiFOXS [284].

Data and the related models were deposited in the SASBDB data base (https://www.sasbdb.org/). The SASBDB data base accession codes and experimental SAXS parameters are reported in Table 7-1.

7.3 Case 2: Structural plasticity enables evolution and innovation of RuBisCO assembies[‡‡]

7.3.1 Introduction

The vast majority of proteins oligomerize into higher order molecular assemblies; however, the phenomenon of protein oligomerization has long remained paradoxical, despite its prevalence in nature. Two contrasting — although not mutually exclusive — modes for the evolution of oligomerization are commonly rationalized. In one, the assembly of a fixed number of subunits is required for protein function (e.g., substrate binding and catalysis), with selection driving the adoption of oligomeric states over time to maintain activity[337,338]. In the other, mutational trends result in a propensity to oligomerize, albeit decoupled from catalytic activity[339-341]. Given that alterations to protein structure enable and/or potentiate new functions, understanding how new oligomeric states originate is a fundamental, yet poorly understood, aspect of protein evolution. Although there has been great interest in elucidating the molecular factors driving new forms of oligomerization, such studies require the comprehensive characterization of entire protein families across time and phylogeny; however, most structural studies have focused on small subsets to single representatives of protein families[340]. Without first order knowledge describing the distribution and diversity of protein oligomerization, we have been largely unable to discern the degree of oligomeric drift that occurs during the evolutionary process and how it may contribute to new functional commitments of proteins.

Rubisco is one such enzyme where biological function is predicated upon oligomeric state but is also capable of adopting multiple assemblies. All rubiscos are composed of a core dimeric scaffold required for catalytic activity; however, complexes from dimeric building blocks can assemble into higher order structures. The vast majority of research has centered on form I rubisco, as the biological source of nearly all organic carbon on Earth, yet the evolutionary events leading to its unique hexadecameric assembly — eight large and eight small subunits — remain elusive[342-344]. In contrast, all other forms of rubisco across the tree of life lack small subunits, and instead assemble as a variety of homomeric complexes. In particular, representatives of form II rubisco have been shown to assemble as either dimers or hexamers, thus offering a unique system in which to study the evolution and transitions of oligomerization of a related enzyme lacking the strict structural requirements of the form I enzyme[345-348]. Here we investigate the diversity and evolutionary trajectory of oligomerization in form II rubisco, revealing an unprecedented level of structural plasticity which underlies the interconversion between, and innovation of, multiple oligomeric states.

7.3.2   *Results*

7.3.2.1   Diversity-driven sampling across extant rubisco reveals complex history of oligomeric state
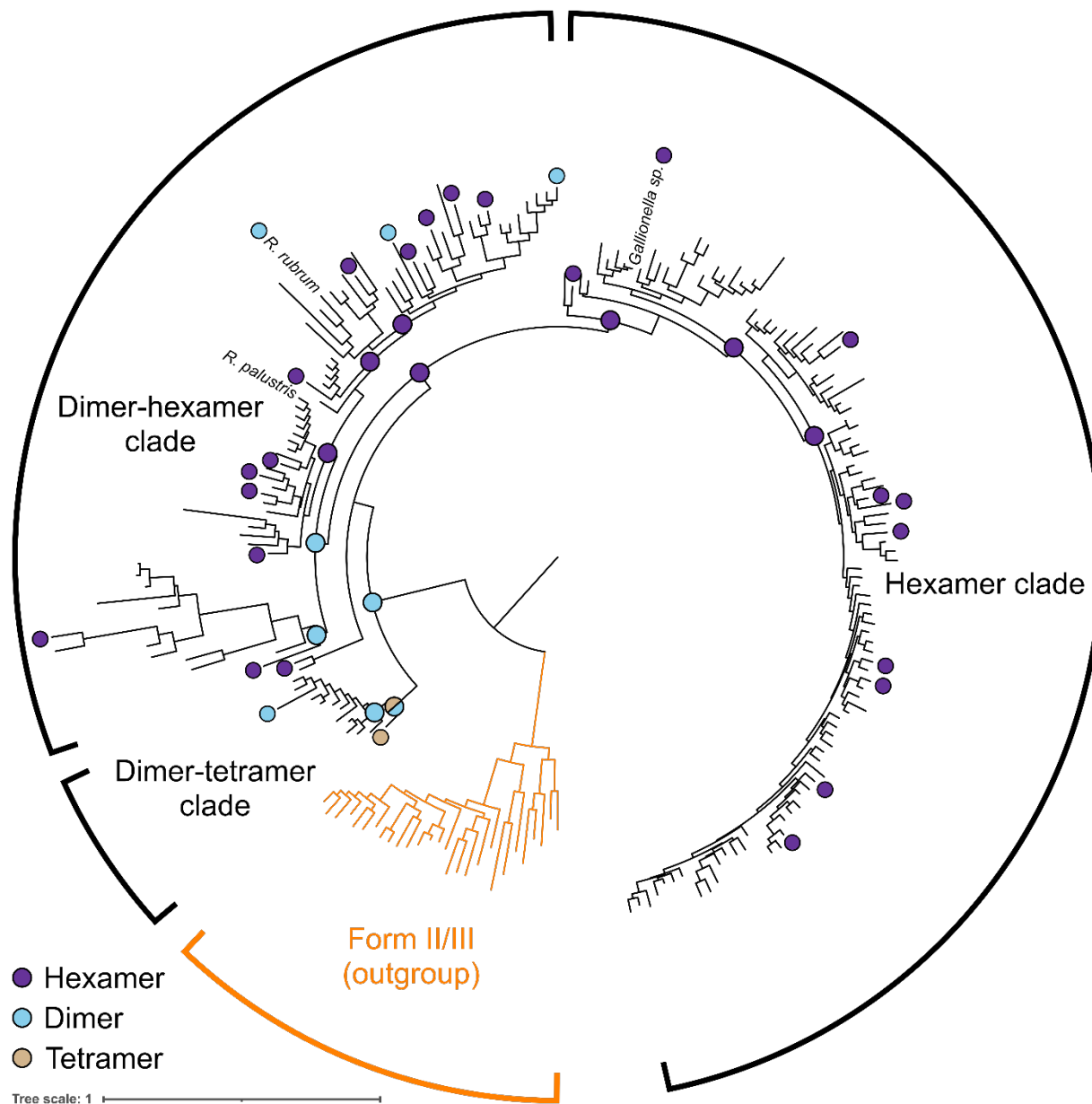
Figure 7-8. Diversity-driven sampling reveals plasticity of RuBisCO oligomeric state. Phylogenetic tree of form II RuBisCO, form II/III serving as outgroup. Selection of presented sequences detailed in Materials and Methods Section 7.3.5. Oligomeric states of characterized extant enzymes are indicated at tips, and those of ancestral enzymes are indicated at corresponding nodes.

To better understand the phylogenetic distribution of oligomeric states found within form II rubisco, we structurally characterized 28 candidates spread across the phylogeny (Figure 7-8, Figure 7-12). From a recent library of form II rubisco, all homologs were heterologously expressed, purified, and analyzed by size exclusion chromatography coupled with small-angle X-ray scattering and multi-angle light scattering (SEC-SAXS-MALS; see Chapter 2)[1,2,9,348]. From the collected SAXS profiles and estimated molecular weights, we observed both dimers and hexamers, with 23 out of 28 adopting the hexameric state (Figure 7-8, Figure 7-13, Table 7-3). Notably, we collected SAXS and MALS data on a novel tetrameric enzyme, representing an entirely new oligomeric state of rubisco that has never been structurally characterized, suggesting an unprecedented level of quaternary diversity within form II rubiscos (Figure 7-13, Table 7-3)[348]. It is commonly believed that form II rubisco exist primarily as dimers; this is largely assumed because the first solved crystal structure of a rubisco was a dimeric form II rubisco from *Rhodospirillum rubrum*[345]. However, more recently the crystal structure of two hexameric structures have also been described[346,347]. By taking a phylogenetic approach to characterizing this entire protein family, we demonstrate that the vast majority are actually hexameric (Figure 7-8). Our findings illustrate the need for diversity-driven studies to correct preconceived biases resulting from sparse structural sampling in our understanding of how protein structures and entire protein families evolve over time.

Upon mapping the characterized oligomeric states onto the phylogeny, we uncovered three distinct patterns of oligomerization representing parallel evolutionary trajectories. One clade, herein referred to as the hexamer clade, is entirely comprised of hexamers, including a previously characterized *Gallionellaceae* enzyme[347] (Figure 7-8). In contrast, the dimer-hexamer clade displays several dimeric enzymes interspersed between hexamers, highlighting the structural plasticity of form II rubiscos in this clade (Figure 7-8). These structural reversions provide a unique case study to demonstrate how the dimer-hexamer clade is not structurally entrenched, and thus has the ability to drift from one state to another. Notably, this clade includes the benchmark form II rubisco from *R. rubrum*, as well as another structurally characterized hexamer from *R. palustris*[345,346]. Finally, the dimer-tetramer clade is composed of dimers and the novel tetrameric rubisco, providing a glimpse into how nature has been able to evolve and innovate novel oligomeric states (Figure 7-8). Overall, our diversity-driven structural characterization across form II rubisco reveals three different clades with three unique evolutionary histories: 1) structural entrenchment, 2) reversible transition states, or 3) innovation of entirely novel oligomeric states.

7.3.2.2 Reconstructing evolutionary trajectories across time elucidates plasticity of oligomeric state

To expand beyond sampling extant sequences, we recapitulated the evolutionary histories of these three different clades by characterizing the ancestral nodes across the form II rubisco phylogeny. We synthesized and characterized 12 ancestral sequence reconstruction enzymes in a manner similar to the extant form II rubiscos. The most recent common ancestor (MRCA) of all form II rubiscos (node 8) was dimeric, reinforcing the most parsimonious scenario of a dimeric origin of form II rubisco (Figure 7-8 and Figure 7-14). Notably, the dimer-tetramer clade MRCA (node 9) adopts both a dimeric and tetrameric state in solution as captured by SEC-SAXS-MALS

(Figure 7-8 and Figure 7-14). The subsequent sister node 10 forms a dimer, representing the origin of the dimers within the dimer-tetramer clade. The biphasic assemblies of node 9 demonstrate the structural plasticity of form II rubisco, as it reprints an evolutionary intermediate that has the propensity to form either a dimer or tetramer prior to the eventual commitment to either trajectory. This evolutionary plasticity is not observable from solely sampling extant enzymes, highlighting the need for ancestral enzyme characterization to visualize oligomeric interconversion within structurally plastic enzyme families.

In conjunction with the oligomeric state of the form II MRCA, analysis of nodes within the dimer-hexamer clade revealed multiple independent interconversion events. From the most ancestral dimer, an intermediary hexamer (node 23) underwent a reversion event resulting in the ancestral dimer preceding the dimer-hexamer clade (node 127) (Figure 7-8 and Figure 7-14). From node 127, the dimer then formed and maintained the hexameric state over several branch points, before reverting once more into extant dimers (Figure 7-8 and Figure 7-14). This clade reinforces the idea that the oligomeric state in some protein families may be quite plastic, allowing for reversions and transitions between different states. This is best demonstrated by a pair of two closely related homologs from *Insolitispirillum peregrinum* and *Rhodospirillaceae* bacterium BRH_c57 (76.3% identity), which form a dimer and hexamer, respectively. This is in contrast with the hexamer clade, whose ancestral enzymes at nodes 24, 27, and 28 were indeed hexameric as well (Figure 7-8 and Figure 7-14). The hexamer clade suggests that there is some biochemical purpose that has entrenched this clade as hexamers, whereas the dimer-hexamer clade is free of those quaternary structure restrictions.

These observations provide insight as to how evolutionary trajectories may affect patterns of oligomerization of phylogenetically related enzymes: entire clades can adhere to a singular oligomeric state, or plasticity can enable free interconversion over time. Although it has been suggested that a ratchet-like evolution of oligomeric state may drive proteins into higher order assemblies mediated by hydrophobic interactions, not all homomeric or heteromeric complexes form via solely hydrophobic patches[341]. Rubisco offers an interesting counterexample where homomeric complexes form via solvent-accessible polar interactions, which underpins the flexibility of oligomeric state in the dimer-hexamer clade. With no known functional constraint between dimers or hexamers, the dimer-hexamer clade appears to have the oligomeric plasticity to explore and interconvert between both states, whereas the hexamer clade has been captured in a sole oligomeric state, likely stemming from an uncharacterized functional pressure. The extent and pervasiveness of proteins that are amenable to this level of quaternary structure freedom may be hard to determine. However, our analyses provide an important case study on how structural plasticity may enable protein drift through both sequence space and oligomeric state while innovating new forms and functions. This could explain the two states observed in the dimer-hexamer clade and the novel tetramer in the dimer-tetramer clade. However, a functional role may still result in oligomeric entrenchment, resulting in the widespread adoption of a singular oligomeric state (e.g., hexamer clade).

7.3.2.3 Evolutionary innovation of a novel tetrameric rubisco that co-opts a unique dimer-dimer interface

Previously, rubisco has only been described to adopt dimeric and cyclic core structures (e.g., hexameric, octameric, and decameric), with previous work suggesting the existence of a tetrameric assembly from the organism *Sulfurivirga caldicuralii* (abbreviated *Sc*)[348]. SEC-SAXS-MALS analysis on the *Sc* enzyme revealed a molecular weight of 218.3 kDa, in agreement with a proposed composition of four large subunits of approximately 50 kDa each (Table 7-3). The collected SAXS curve did not match trends observed from either dimeric or hexameric rubisco, further suggesting that the tetrameric state is distinct from other form II structures (Figure 7-13). Additionally, the SAXS curve did not fit a tetrameric structure generated by removing two dimers from the octameric core of a form I rubisco, thus informing us that the assembly of the *Sc* tetramer is distinct from that found within the octameric rubisco.

To better understand the oligomeric state of *Sc* rubisco, we solved its crystal structure at 1.7 Å resolution, clearly displaying its tetrameric assembly (Figure 7-9A). The arrangement of the pair of dimers precludes the formation of a central solvent channel, a feature observed in all oligomers of rubisco that form a cyclic structure. Identification of interface residues revealed a compacted interface aligned more closely to the center of each dimer and distinct from that of the hexamer's (Figure 7-9B and Figure 7-15). This illustrates the means by which novel oligomeric states can be innovated over the course of structural drift, as the tetramer is differentiated both phylogenetically and structurally from the hexamer, thus precluding the use of the larger oligomeric state as the template. In conjunction with our phylogenetic analyses, this observation highlights the unique assembly of the tetramer, as its early divergence from the remainder of form II rubisco precedes the innovation of the hexameric state yet remains maintained after the divergence from ancestral node 9 into the remainder of the dimer-tetramer clade.

Moreover, when compared to the octameric cores of form I and I' assemblies, it becomes apparent that the combination of two tetramers would not yield a conventional octamer (Figure 7-9C). It can be purported that a rubisco octamer (a tetramer of functional dimers) could be assembled from two tetramers (dimers of dimers), in accordance with our understanding of oligomeric assembly[98 and Evolution of Protein Complexes,349]. However, dimers within an octameric core are vertically aligned in parallel, whereas the *Sc* rubisco's central axis results in the observed angled assembly. This further illustrates the differences between the evolutionary trajectory of form I and form II oligomeric state; while interface mutations in an octamer could result in the formation of two tetramers, the geometric differences between a pair of form I dimers and the form II tetramer preclude this possibility and instead suggest the independent innovation of the tetrameric state.

Structural plasticity has been proposed to affect the oligomeric state of enzyme families in two distinct ways: 1) large geometric changes can be buffered by plasticity and result in the maintenance of oligomeric state or 2) plasticity can underpin geometric flexibility and give rise to multiple oligomeric states[350]. Form I rubisco may represent an example of the former situation, as it remains highly constrained by its base octameric assembly, thus resulting in minor changes to the angle of dimers within the octamer without changes in the entirety of its oligomeric state. In contrast, we demonstrate that form II rubisco falls into the latter category, where a highly plastic ancestral dimer may have fortuitously bound a second dimer and gave rise

to tetramerization, while subsequent evolution of singular dimers produced the precursor interfaces necessary for hexamerization. Ultimately, the tetrameric form of rubisco exemplifies how the structural plasticity of proteins enables the innovation of entirely novel oligomeric states through the recruitment of novel surface residues to mediate protein-protein interactions.
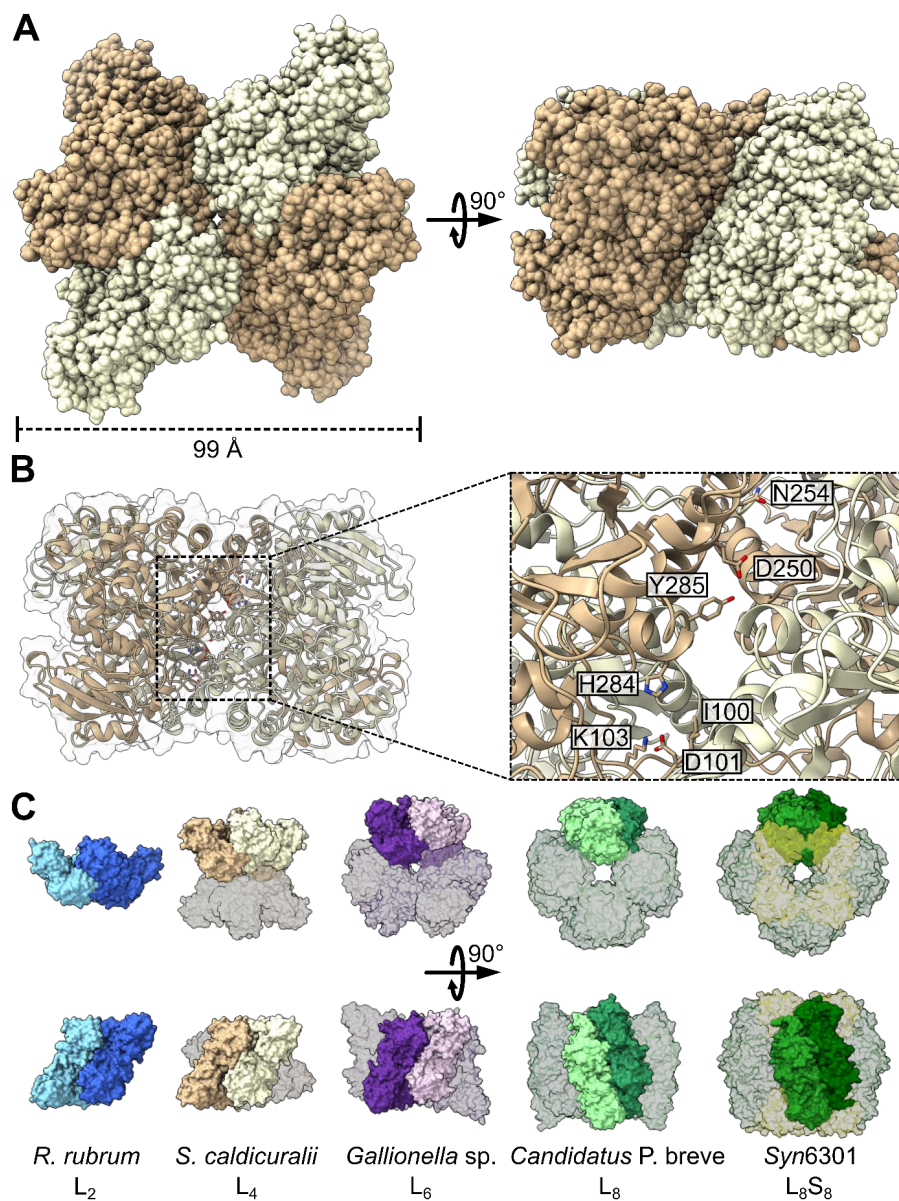
Figure 7-9. Crystal structure of a tetrameric RuBisCO. (A) Structure of S. caldicuralii RuBisCO resolved at 1.7 Å. (B) Interface cutaway of S. caldicuralii tetramer with candidate residues indicated. (C) Comparison of RuBisCO oligomeric states illustrating dimer positioning within a multimer. Form II dimer, tetramer, and hexamer are shown alongside form I′ octamer and form I hexadecamer. Protein Data Bank (PDB) codes (left to right): 5RUB, 7T1C, 5C2C, 6URA, and 1RBL.
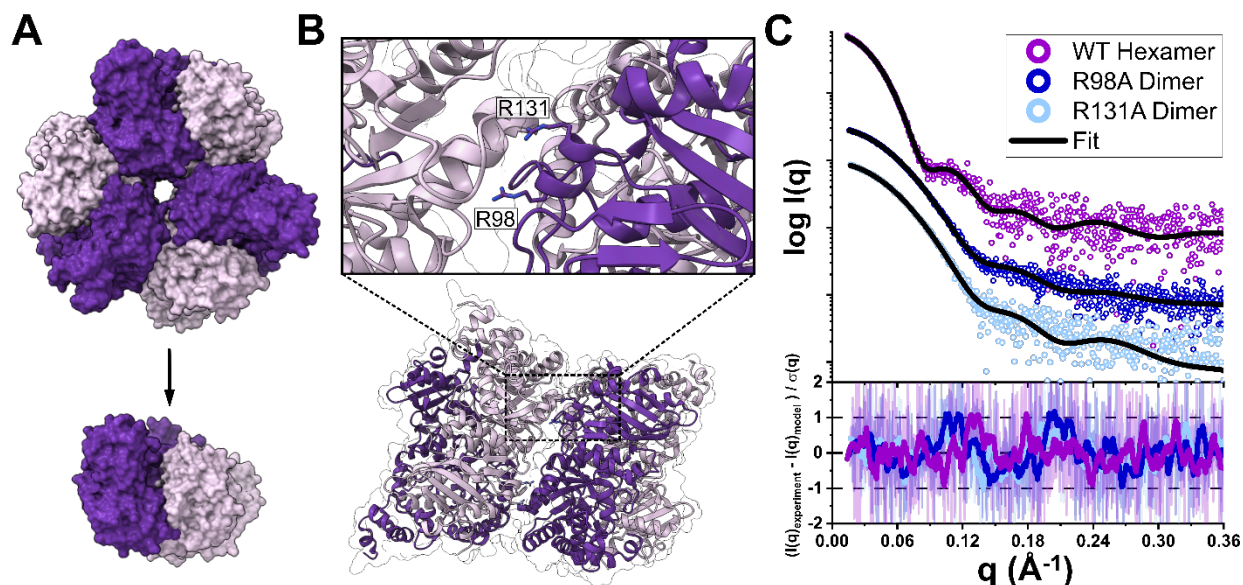
Figure 7-10. Hexamers can readily form dimers through mutations of residues coordinating the interdimer interface. (A) Modeling disruptions at the interdimer interface of the hexameric Gallionella sp. structure (PDB: 5C2C) to shift its oligomeric state from hexamer to dimer. (B) Interface cutaway indicating candidate residues. (C) SAXS curves of experimental data for wild-type (WT) enzyme, R98A, and R131A mutants and theoretical fit models for hexameric and dimeric states (PDB: 5C2C and 5RUB, respectively). Fit residuals shown below.

7.3.2.4   Structural plasticity enables reversions to simpler oligomeric states

To investigate the hypothesis that molecular complexes are subject to ratchet-like evolution that entrenches oligomeric states of proteins[341], we tested how easily form II proteins could revert from higher order hexamers to the simpler dimeric state. To identify the specific interface residues involved in higher-order assembly of rubisco, we utilized Protein Contacts Atlas to analyze the interdimer interface of a previously characterized hexameric *Gallionellaceae* enzyme (Figure 7-10A)[347,351]. From a list of computed atomic interactions, we discovered two arginine residues at positions 98 and 131 capable of forming multiple interactions across the interface, including a potential salt bridge with an aspartic acid residue at position 256 (Figure 7-10B and Figure 7-16). Using the *Gallionella* sp. enzyme as a template, we conducted sequence conservation analysis to further analyze the composition and maintenance of the hexameric rubisco interface (Figure 7-17A). Across all extant hexamers identified from our characterization experiments, the R98 residue proved to be more conserved than R131, though neither proved to be especially variable in comparison to a less conserved residue, such as Y358 (Figure 7-17B). However, when comparing patterns within clades, the residue identity of position 131 is highly variable in the dimer-hexamer clade compared to the hexamer clade, wherein both R98 and R131 are highly conserved (Figure 7-17C-D). The variability in interface residue conservation across clades further demonstrates the mechanisms of differentiation between the dimer-hexamer and hexamer clades, as R131 may serve as one such residue that strengthens the hexameric state

within the hexamer clade, whereas the plasticity within the dimer-hexamer clade resulted in more variable identities at that same position.

To query the contribution of R98 and R131 to the maintenance and stability of the inter-dimer interface, we conducted alanine substitutions at both positions and characterized the point mutant enzymes in the same manner as their wild-type counterpart. Strikingly, both the R98A and R131A mutants adopted the dimeric state, as verified by SEC-SAXS-MALS (Figure 7-10C). Analysis by protein thermal shift assays revealed a decrease in thermal stability for both mutant dimers relative to the wild-type, with R98A fully denaturing at 10.5°C lower than the WT, and R131A 12.5°C lower (Figure 7-18). These findings are contrary to conventional perspectives on the strength and maintenance of oligomeric state, as a single residue substitution resulted in loss of a higher-order assembly, though it remained structurally viable in its base state as opposed to an anticipated critical destabilization of the entire enzyme[341,352,353]. Mutational ratchet-based oligomerization is considered irreversible due to the nature of its mechanism, as a disadvantageous property is thought to be conferred to composite subunits were they to be isolated from one another. However, we demonstrate that exposure of the buried hexameric inter-dimer interface does not result in catastrophic destabilization of the enzyme, suggesting that the irreversibility of higher-order oligomerization may be overruled by highly plastic evolutionary trajectories that enable interconversion events akin to our experiments.

7.3.2.5   Oligomerization tunes rubisco activity and kinetic parameters

Although the residues defining rubisco dimer-dimer assembly are distal to the active site, we hypothesized that minor perturbations to the core dimer that still result in drastic changes in quaternary state may affect the kinetic parameters of the enzyme. It has been previously demonstrated that such distal mutations can affect the enzymatic properties of a wide variety of enzymes[354,355]; thus, we investigated the specific implications of oligomeric disruption on rubisco catalysis. We measured the kinetic parameters of the two mutant R98A and R131A enzymes (Table 7-2). Due to rubisco's dual carboxylase and oxygenase activities, measured parameters include turnover numbers, (k_{cat}^C, k_{cat}^O, respectively), Michaelis constants for CO_2 and O_2 (K_C, K_O), and the rubisco specificity factor ($S_{C/O}$). Both mutants displayed decreased k_{cat}^C, by approximately 30% for R98A and approximately 22% for R131A relative to the wild type (Table 7-2). However, the mutant enzymes displayed an increase in specificity factor, with $S_{C/O}$ values approximately 1.17 times higher in R98A and 1.13 times higher in R131A than the wild type (Table 7-2). In light of the modest changes to K_C and k_{cat}^O, the change in specificity appears to be largely driven by a dramatically decreased affinity for oxygen as a substrate, with K_O values being approximately 1.68 times higher in R98A and 2.15 times higher in R131A than the wild type (Table 7-2).

While the rubisco inter-dimer interface is distinct from the active site, it was previously unknown whether rubisco activity was affected by the loss of higher-order assemblies. These experiments demonstrate that catalytic activity is indeed maintained in the absence of the wild-type quaternary structure. In comparison to existing kinetic measurements for other form II rubisco, it is also important to note that despite exhibiting decreased k_{cat}^C values of the R98A and R131A mutants, both are still extremely high values, ranking within the top seven fastest rubisco ever studied, with the wild type *Gallionella* sp. enzyme as the third fastest form II enzyme ever

measured[348]. In light of these considerations, the observation that complete reversion of the hexameric state to the dimeric state resulted in relatively minimal changes to most kinetic parameters is of great interest, as this suggests that the innovation of oligomeric states within the form II evolutionary trajectory may have incurred minimal functional penalty.

7.3.2.6 Engineering increased oligomeric complexity

To further query the structural plasticity of form II rubisco, we tested how readily we could introduce surface mutations to the enzyme in order to engineer higher order assemblies of rubisco from the base dimer. We developed a Rosetta-based computational pipeline to model the transition of a dimer to a hexamer, dubbed "2-to-6". Two closely related rubiscos from the dimer-hexamer clade were used as a template hexamer and a candidate dimer, where the dimer (*Insolitispirillum peregrinum*, hereon referred to as *Ip*) and the hexamer (*Rhodospirillaceae* bacterium BRH_c57, hereon referred to as BRH_c57) share 76.3% sequence identity (Figure 7-19A). Additionally, we solved the crystal structure of the BRH_c57 hexamer to identify the residues participating in its inter-dimer interface, in conjunction with Rosetta modelling of a mutant "2-to-6" *Ip* hexamer (Figure 7-19A-B). Initially, simple mutational experiments were performed using the interface interactions derived from the BRH_c57 structure, though these did not result in an increase in oligomeric state. Thus, we utilized a more rigorous modeling and scoring protocol within Rosetta to screen 128 combinations of different interface residue mutants, with a total of seven residue substitutions (K98R, A134R, T148R, G151E, G281Q, T282Q, G358Q) introduced into the *Ip* sequence based on the top candidate (Figure 7-11B). The candidate 2-to-6 sequence was then expressed, purified, and characterized by SEC-SAXS-MALS, confirming the generation of a hexameric *Ip* rubisco. Of the seven substitutions, the R98, R148 and Q282 residue identities were also present in the hexameric sequence conservation analysis conducted previously, while the remainder were unique to the BRH_c57 enzyme. Notably, the G358Q mutation was predicted to position R134 and enable an interaction with E151 — an interaction not observed in the original BRH_c57 interface (Figure 7-19C). Our engineered protein demonstrates how higher oligomeric states can be assembled through point mutations at the interdimeric interface, with further structural differentiation innovated by residue positioning.

Further analysis of the SEC-SAXS-MALS sample revealed an unexpected biphasic population of the engineered 2-to-6 enzyme. A hexameric assembly was indeed captured and verified by comparison of its SAXS scattering data with the BRH_c57 hexamer, but a second dimeric state was also present in the purified sample (Figure 7-11C). The presence of both oligomeric states is akin to the behavior exhibited by the dimeric/tetrameric ancestral node 9, suggesting the capture of an intermediary transitional state prior to commitment to either the dimeric or hexameric state. While the 2-to-6 sequence ranked highest from Rosetta modeling, the distribution of dimeric and hexameric species in the experimental sample suggests that residues not involved in hydrogen bonding at the inter-dimer interface may play a key role in enabling the complete transition to a hexameric state, in agreement with previous observations regarding the role of distant mutations in oligomerization (Figure 7-20)[350]. Our results demonstrate how a small number of residues — only seven mutations — can enable an increase in oligomeric state, providing insight into the requisite degree of plasticity necessary for innovation of larger oligomeric states. Although

previous work has relied on the introduction of hydrophobic patches to enable the self-assembly of large protein complexes[356], we demonstrate that subtle structural differences of polar amino acids on a solvent-exposed surface can be utilized to dictate the proper formation of predicted interactions constituting protein-protein interfaces.

Table 7-2. Dimers formed from hexamers demonstrate how distal mutations from the active site mediate enzymatic tradeoffs and fine tune kinetic properties of rubisco. Values are means ± S.E. with n indicated in brackets.

Rubisco	Oligomeric state	k_{cat}^C (s^{-1})	K_C (µM)	$S_{C/O}$	k_{cat}^O (s^{-1})	K_O (µM)
Gallionella sp. WT	L$_6$	15.7 ± 0.9 (5)	172 ± 29 (5)	22.0 ± 1.3 (5)	0.38	92 ± 15 (4)
R98A	L$_2$	11.1 ± 1.2 (4)	170 ± 25 (4)	25.7 ± 1.8 (6)	0.39	155 ± 16 (4)
R131A	L$_2$	12.3 ± 0.9 (5)	198 ± 12 (4)	24.9 ± 0.9 (6)	0.50	198 ± 21 (4)
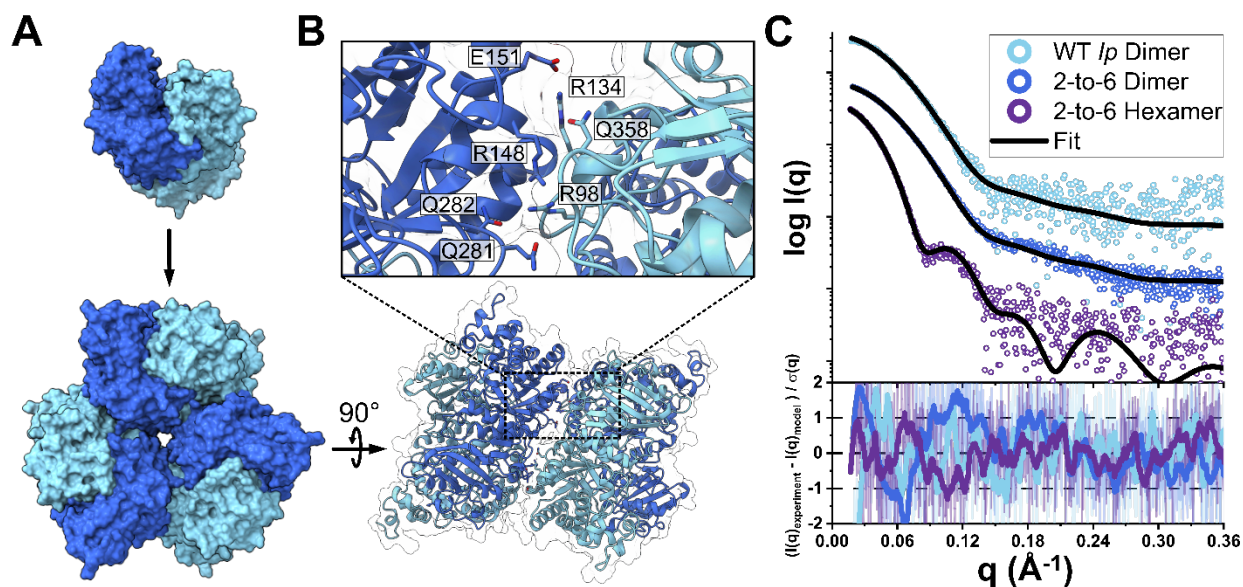
Figure 7-11. Structurally guided engineering recapitulates dimer-to-hexamer oligomeric transition. (A) Modeling of the interdimer interface to convert the dimeric I. peregrinum RuBisCO into a hexamer. (B) Interface cutaway of introduced mutations in the hexameric I. peregrinum homology model to engineer a network of side chain interactions to mediate an oligomeric shift to hexamerization. (C) SAXS curves of experimental data for wild-type and engineered I. peregrinum enzyme and theoretical fit models for both hexameric and dimeric states present in the same characterized sample [PDB: 7T1J and homology model of I. peregrinum (Ip) dimer, respectively]. Fit residuals are shown.

7.3.3 Discussion

Our understanding of the diversity, origins, and trajectories of protein oligomeric states has been largely incomplete due to the lack of diversity-driven studies required to properly assess how quaternary structure evolves over time. Our findings reveal an unprecedented level of structural plasticity underlying an assortment of unique evolutionary trajectories within a single protein family, ranging from structural entrenchment, interconversions, and innovation of new oligomeric states. The characterization of a novel tetrameric form of rubisco best highlights how evolution continually explores sequence space and co-opts surface residues in the formation of entirely new oligomeric states. Our findings demonstrate how quaternary structure may not be as rigid and constrained as previously assumed. Because the majority of proteins form molecular complexes, the underappreciation of this higher form of structural plasticity may have larger implications on many other protein families, where quaternary structure can play a key role in drug targets[357,358], human diseases[359,360], and general function[361-363].

Oligomerization remains critical for the control of biological processes, yet the origins and evolution of the diversity of oligomeric states observed across nature have been poorly understood[364]. It has been recently hypothesized that proteins increase in oligomeric complexity due to ratchet-like evolution mediated by hydrophobic interactions[341]. Although there are examples of this, not all molecular complexes are formed and stabilized through hydrophobic patches, as we have demonstrated in this instance via the solvent-accessible polar interactions found in rubisco. The observed plasticity of form II rubisco illustrates the prevalence of oligomeric interconversion events in nature, demonstrating how evolutionary intermediaries can drift between two distinct assemblies prior to the evolutionary accumulation of additional mutations that result in commitment to either assembly. In the absence of strong selective pressures, the mutations that resulted in structural differentiation can be reversed, accordingly generating an overall reversion of oligomeric state. However, the presence of functional pressures can further select for and entrench a particular oligomeric state, thus precluding any further reversion events. Conservation of interface residues reveals the mechanism by which these states exhibit these patterns of oligomerization, as a highly conserved set of interface residues may be found across all extant multimers, albeit bolstered with additional stabilizing contacts in clades demonstrating a strong commitment to a singular oligomeric state.

Our findings on form II rubisco provide the requisite evolutionary reference point to understand the evolutionary trajectory and structural basis of form I rubisco, the most abundant enzyme on our planet. Unlike all other forms of rubisco, the distinguishing feature of form I rubisco is its unique incorporation of small subunits to assemble its iconic heteromeric complex composed of eight large and eight small subunits. Form I rubisco likely underwent an early differentiation event from an ancestral dimeric state of all rubisco, which subsequently strongly entrenched the octameric core assembly with the acquisition of the small subunit. While the initial binding event between an ancestral octamer and a small subunit-like protein may have occurred with no tangible benefit conferred to either protein (i.e., via constructive neutral evolution), extant form I enzymes suffer from drastically decreased activity in the absence of their native small subunits, thus predicating overall activity on the hexadecameric assembly[365]. However, form II rubisco do not demonstrate a noticeable trend relating oligomeric state to carboxylation activity, further suggesting that the function of form II rubisco is largely independent of its oligomeric state

(Figure 7-21). Overall, the comparison of the two divergent evolutionary paths taken by form II versus form I rubisco provides a dichotomy in structural plasticity versus entrenchment, respectively. The structural plasticity of form II rubisco has resulted in a complex history of various oligomerizations, whereas the innovation and incorporation of the small subunit was the crux in the ratchet-like evolution that gave rise to the form I clade. The strict requirements of form I assembly for catalytic activity are not shared by form II, thus permitting the structural plasticity that enabled the innovation and maintenance of novel oligomeric states.

We have lacked a strong understanding of protein evolution at a structural level, because the traditionally low throughput nature of structural studies has resulted in a sparse phylogenetic sampling of protein families; thus, most of our knowledge largely stems from single to few representatives. Even with advances in protein structure prediction, quaternary structure remains challenging to accurately reconstruct. Thus, diversity-driven studies will help shed light on the complex range of evolutionary paths and disparate oligomeric states that can be observed within individual protein subfamilies. Our results demonstrate how quaternary structure may be inherently malleable until functional roles entrench specific oligomeric states, thus allowing proteins to sample and explore not just sequence space but also disparate oligomeric states. Notably, we also show how changes in quaternary structure may also contribute to the tuning of enzyme kinetics, providing a potential avenue of selective pressure on oligomeric state. Given the central role oligomeric state may play in many proteins, it remains to be shown how prevalent quaternary structural plasticity is across nature, as it may represent a nuanced, yet important, contributor shaping the evolution of protein structure and function.

7.3.4   Supplemental Materials

7.3.4.1   Supplementary Figures
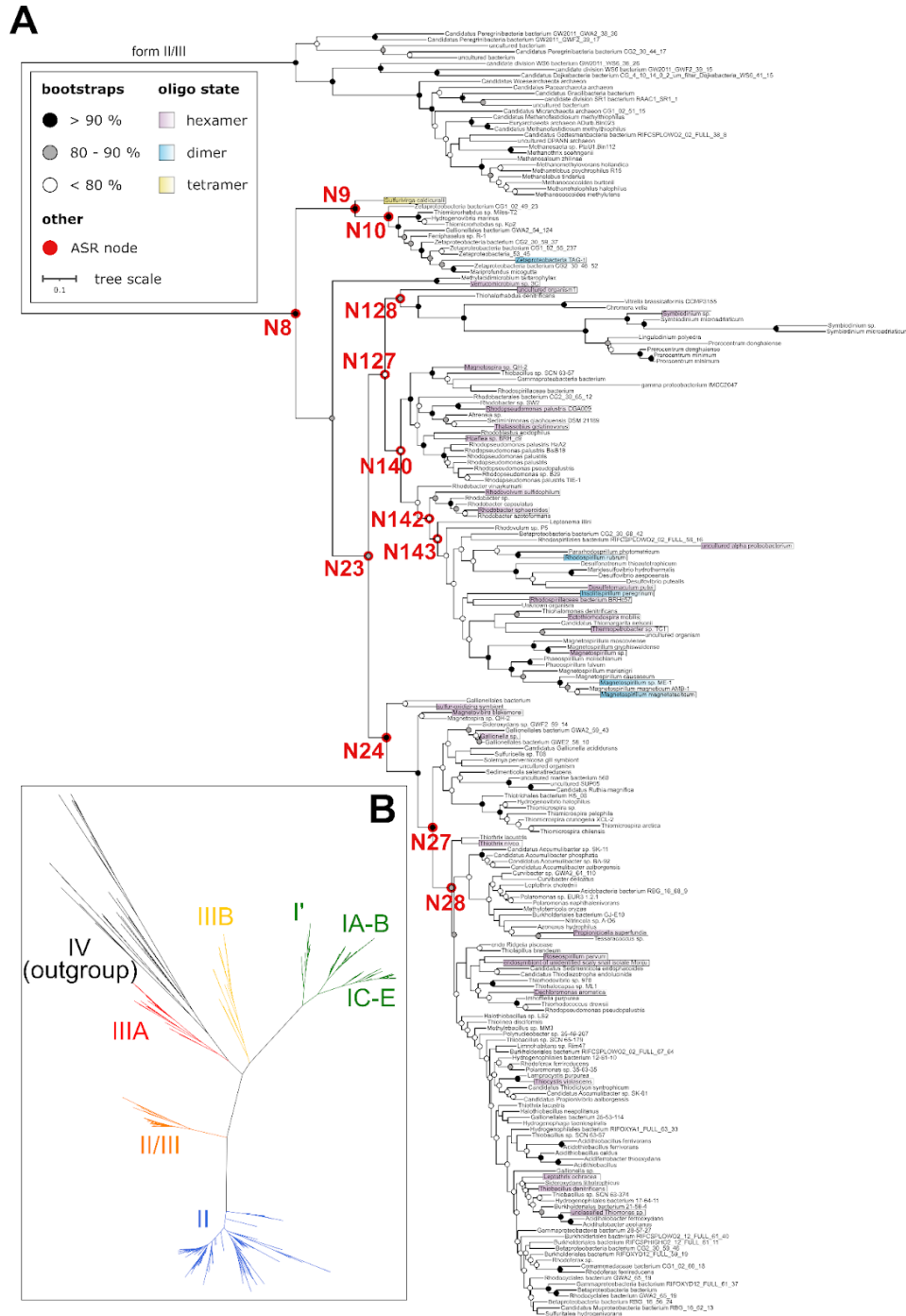
Figure 7-12. Expanded phylogenetic tree of form II rubisco and all rubisco forms. (A) Form II rubisco phylogeny shown with species names at tips. Known oligomeric states indicated in color. Ancestral sequence nodes highlighted and labeled in red. (B) Phylogenetic tree of all rubisco forms. Clades color-coded with accompanying names.
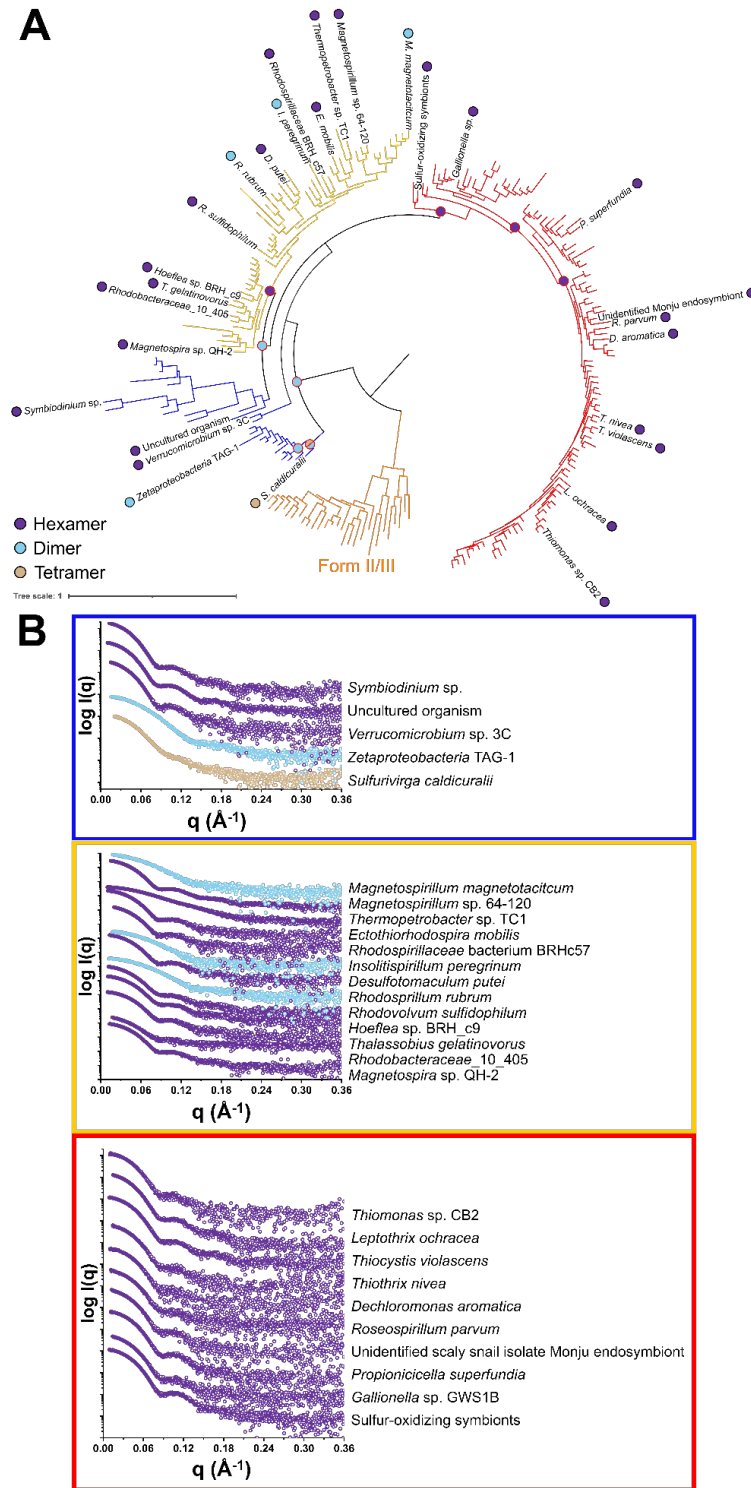
Figure 7-13. SAXS curves of 28 analyzed form II rubisco. Color-coded regions of phylogenetic tree correspond to indicated SAXS curves.
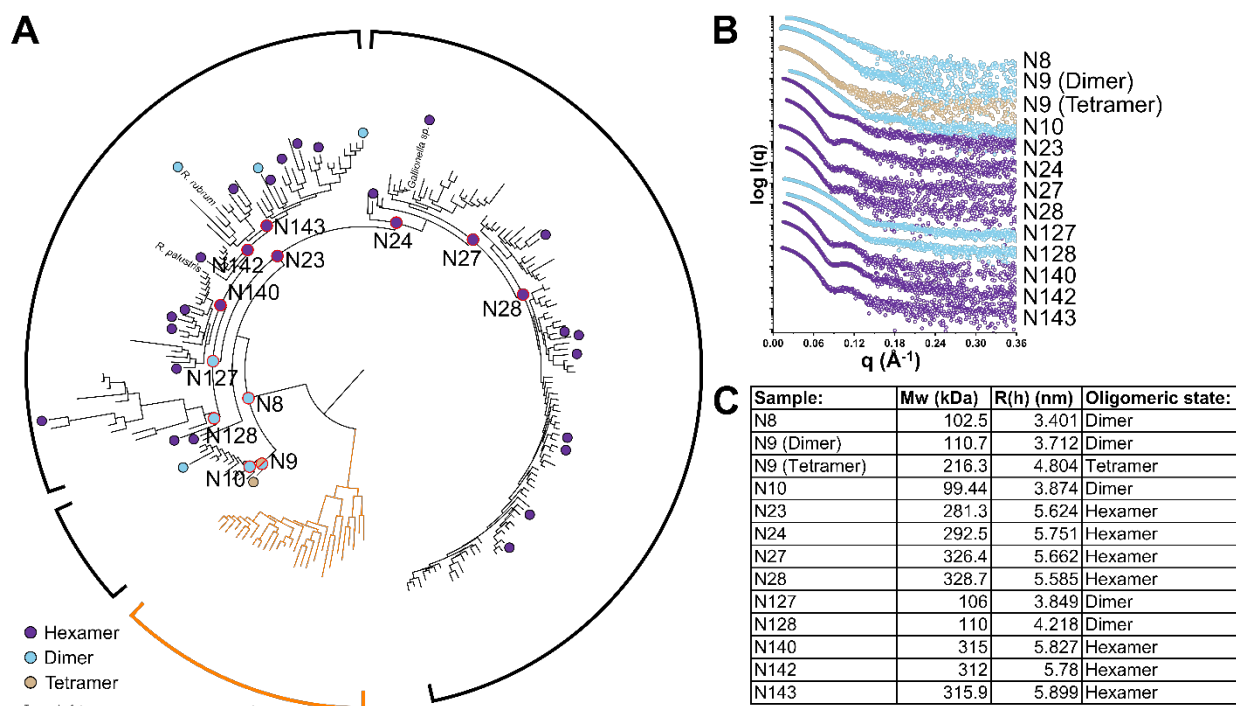
Figure 7-14. SEC-SAXS-MALS analysis of ancestral form II enzymes. (A) Form II phylogenetic tree with indicated locations and names of ancestral nodes. (B) SAXS curves of characterized ancestral enzymes. (C) Collected MALS values with measured molecular weights and radii of hydration.

Sample:	Mw (kDa)	R(h) (nm)	Oligomeric state:
N8	102.5	3.401	Dimer
N9 (Dimer)	110.7	3.712	Dimer
N9 (Tetramer)	216.3	4.804	Tetramer
N10	99.44	3.874	Dimer
N23	281.3	5.624	Hexamer
N24	292.5	5.751	Hexamer
N27	326.4	5.662	Hexamer
N28	328.7	5.585	Hexamer
N127	106	3.849	Dimer
N128	110	4.218	Dimer
N140	315	5.827	Hexamer
N142	312	5.78	Hexamer
N143	315.9	5.899	Hexamer
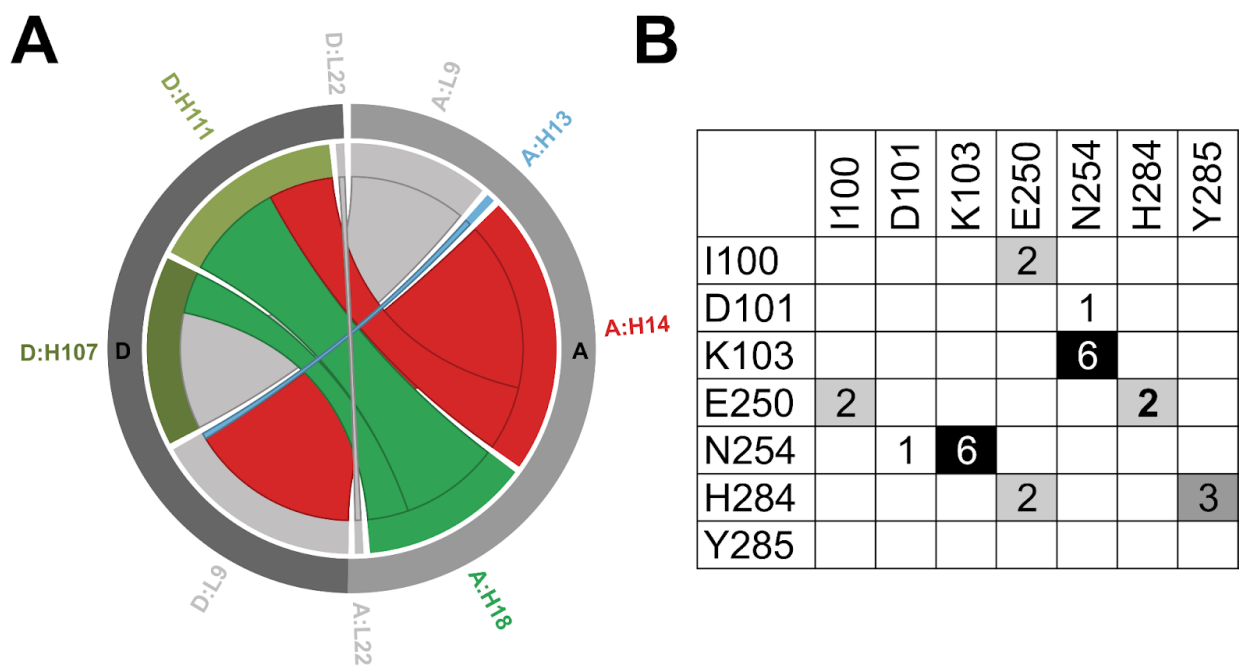
Figure 7-15. Interface residue identification of *S. caldicuralii* tetramer. (A) Protein Contacts Atlas chord plot illustrating interaction network between chains A and D of structure. (B) Heat map of unique side chain-side chain interactions between interface residues.

A

B

	R92	T95	R98	R131	Q134	L135	R148	I149	E249	P253	D254	A255	D256	K257	S281	Q282	Y358	G359	K361
R92										4									
T95																			
R98												1	5		2	7			
R131								2				1	8	6					
Q134						1	2						1						
L135													4						
R148								2									2	1	
I149				1	2														
E249		1																	
P253	5																		
D254					1														
A255			2																
D256			4	6	1	4													
K257				8															
S281			2																
Q282			6																
Y358							2												
G359							1												
K361																			2
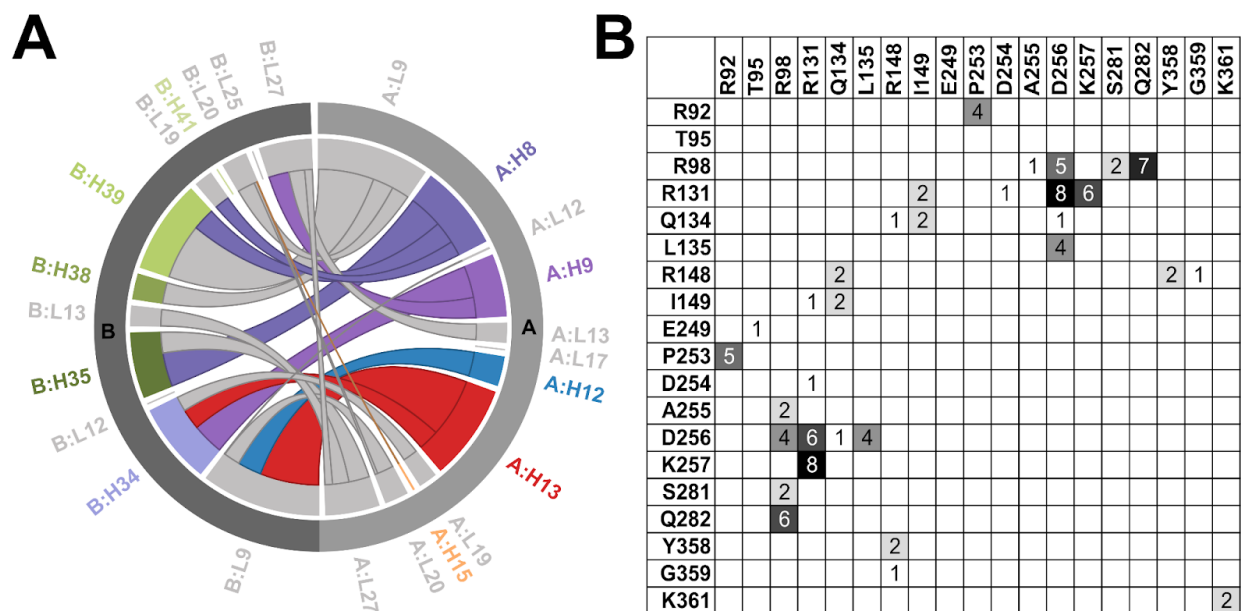
Figure 7-16. Interface residue identification of *Gallionella* sp. hexamer. (A) Protein Contacts Atlas chord plot illustrating interaction network between secondary structures on chains A and B of 5C2C structure. (B) Heat map of unique side chain-side chain interactions between interface residues.
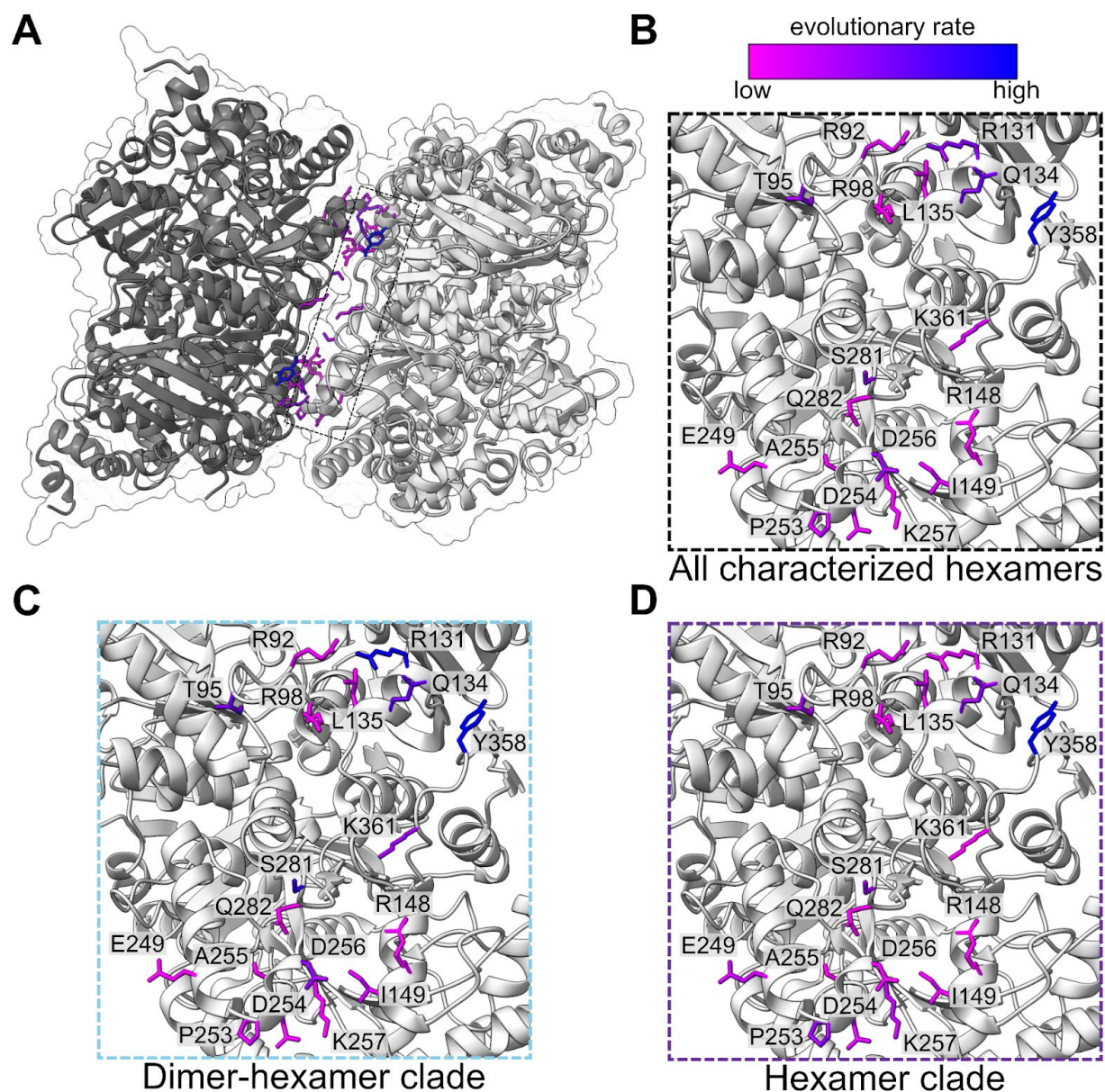
Figure 7-17. Sequence conservation analysis of hexameric *Gallionella* sp. interface. (A) Location of interface residues on *Gallionella* sp. crystal structure (PDB: 5C2C). (B) Conservation of residues as determined from all characterized hexamers. Low evolutionary rate (highly conserved) in magenta, scaling to high evolutionary rate (more variable) in blue. (C) Residue conservation analysis conducted exclusively with hexamers in dimer-hexamer clade. (D) Residue conservation analysis conducted exclusively with hexamers from hexamer clade.
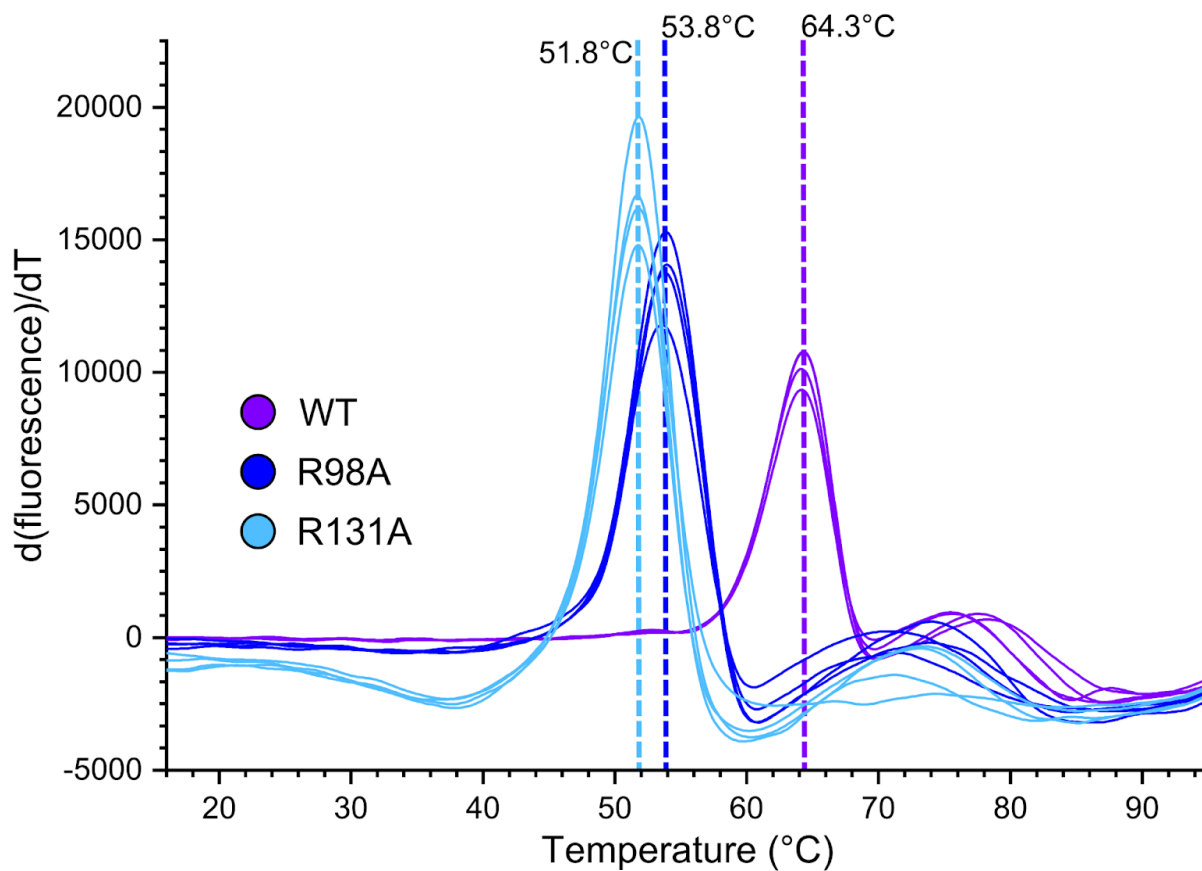
Figure 7-18. Protein Thermal Shift assay melt curves for WT *Gallionella* sp. hexamer and R98A, R131A dimers. Reported T_m values represent the average from four technical replicates.
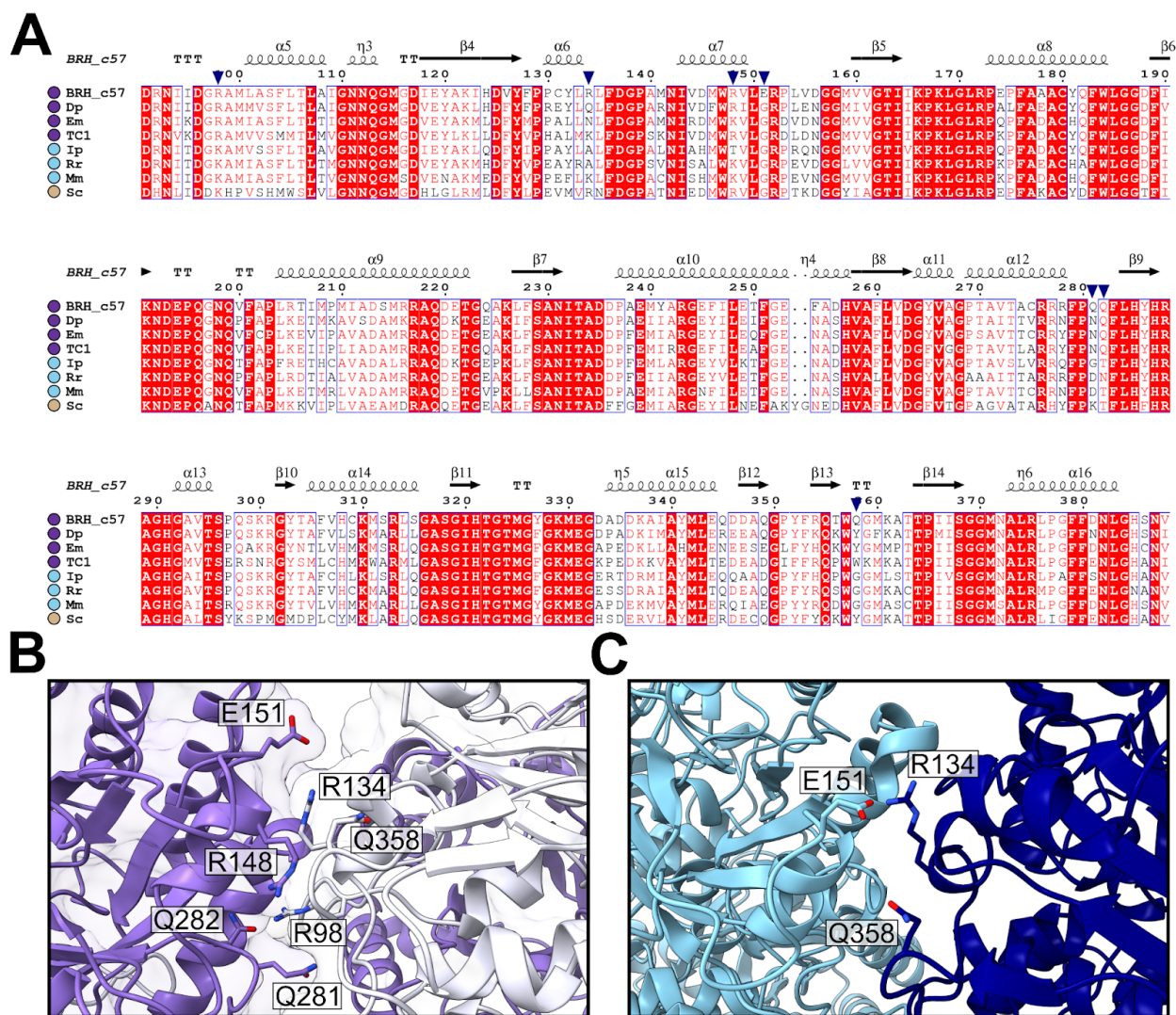
Figure 7-19. Identification and selection of hexameric residues for engineering. (A) Alignment of representative hexamers, dimers, and tetramer. Oligomeric state indicated by circles to left of species name (hexamers; purple, dimers; blue, tetramer; tan). Interface residues indicated by inverted blue triangles. Species listed from top to bottom: *Rhodospirillaceae* bacterium BRH_c57, *Desulfotomaculum putei*, *Ectothiorhodospira mobilis*, *Thermopetrobacter* sp. TC1, *Insolitispirillum peregrinum*, *Rhodospirillum rubrum*, *Magnetospirillum magnetotacitcum*, *Sulfurivirga caldicuralii*. (B) Dimer-dimer interface cutaway of BRH_c57 crystal structure. Residues indicated. (C) Novel inter-dimer interaction formed in hexameric *Ip* mutant.
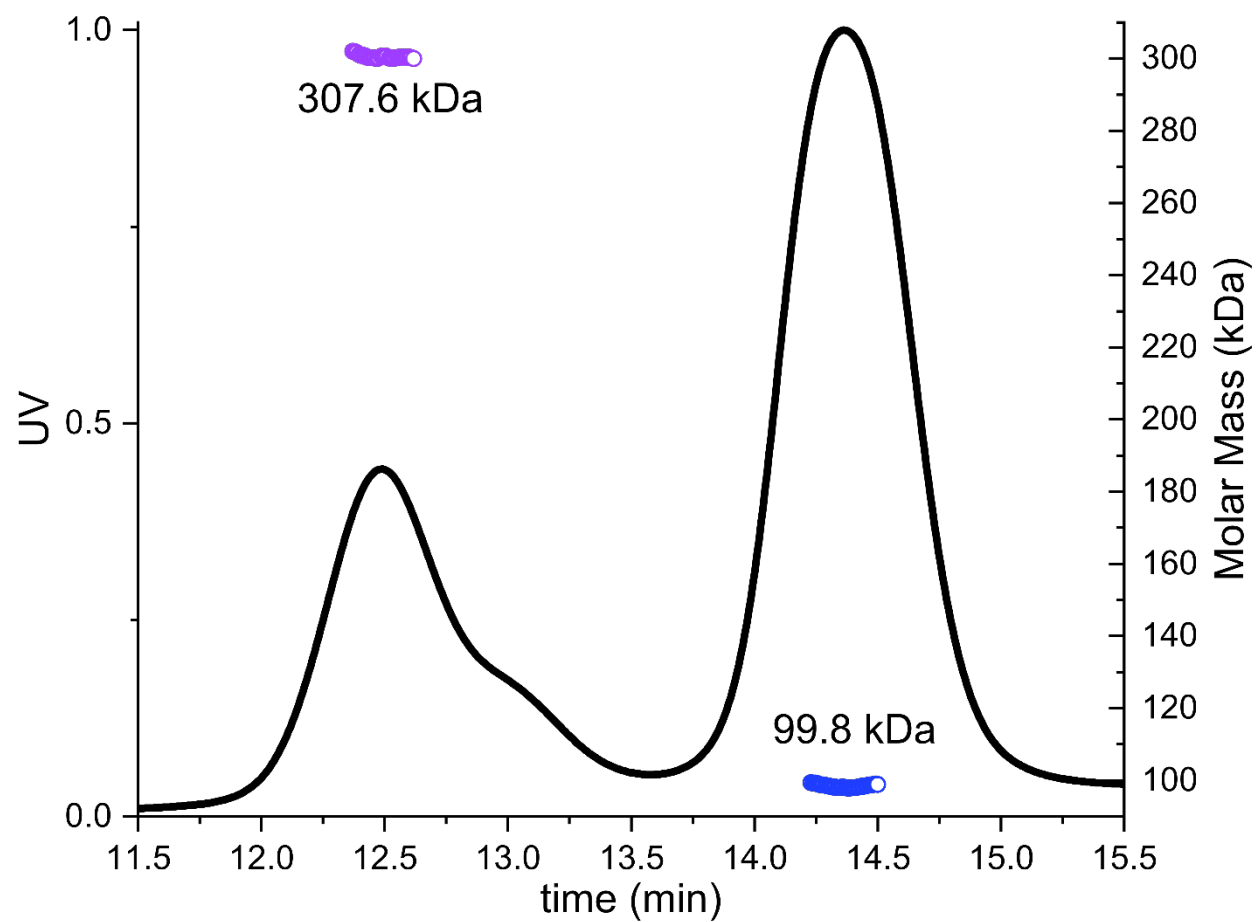
Figure 7-20. SEC-MALS chromatogram of engineered 2-to-6 sample. The molecular weight of each species indicated with each peak.
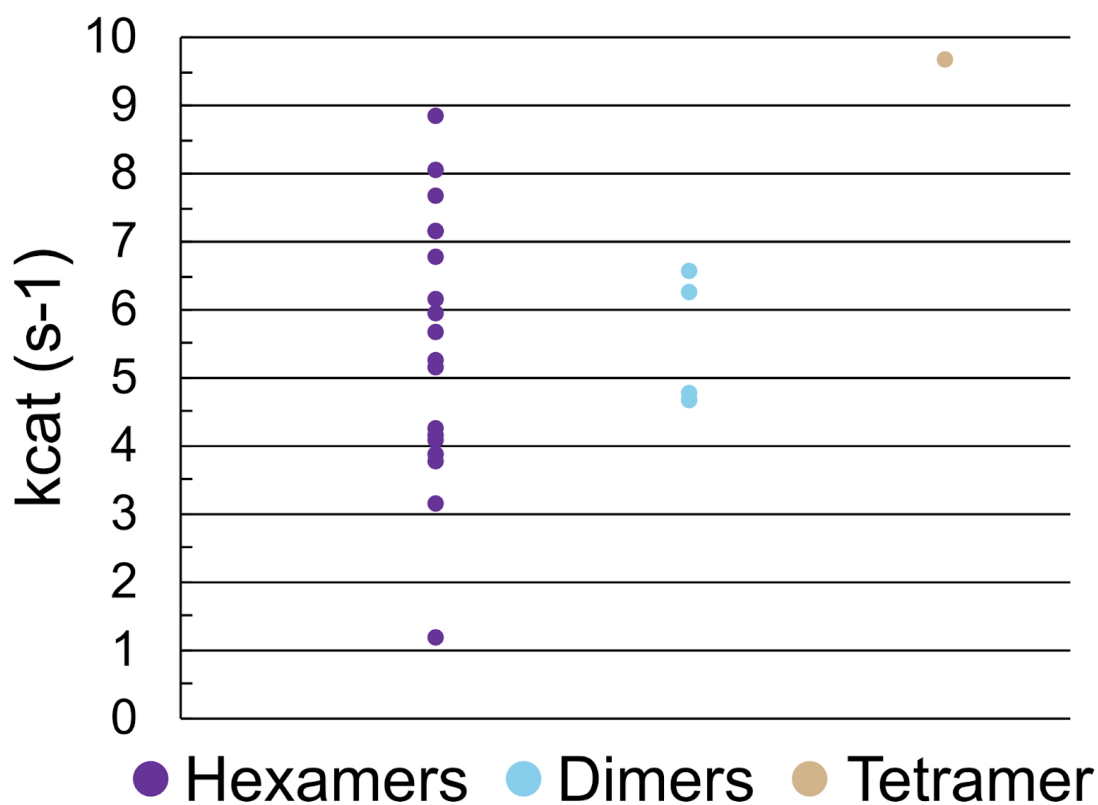
Figure 7-21. Spectroscopic k_{cat}^C values of extant form II enzymes characterized in this study. k_{cat}^C values from [348].

7.3.4.2 Supplementary Tables

Table 7-3. MALS-determined molecular weights and radius of hydration values for 28 characterized form II rubisco.

Sample:	Mw (kDa)	R(h) (nm)	Oligomeric state:
Thiocystis violascens	325.8	5.375	Hexamer
Uncultured organism	332.4	5.645	Hexamer
Leptothrix ochracea	349.4	5.905	Hexamer
Zetaproteobacteria TAG-1	107.2	3.462	Dimer
Thalassobius gelatinovorus	313.2	5.35	Hexamer
Magnetospirillum magnetotacitcum	106	3.727	Dimer
Rhodovolvum sulfidophilum	300.5	5.809	Hexamer
Propionicicella superfundia	317.1	5.25	Hexamer
Thiothrix nivea	314.3	5.404	Hexamer
Verrucomicrobium sp. 3C	319.7	5.616	Hexamer
Thermopetrobacter sp. TC1	314.2	4.86	Hexamer
Symbiodinium sp.	302.4	5.334	Hexamer
Unidentified scaly snail isolate Monju endosymbiont	302.4	5.636	Hexamer
Magnetospira sp. QH-2	299.1	6.531	Hexamer
Magnetospirillum sp. 64-120	325	5.777	Hexamer
Desulfotomaculum putei	324.8	6	Hexamer
Hoeflea sp. BRH_c9	313.2	5.451	Hexamer
Thiomonas sp. CB2	313.7	4.693	Hexamer
Sulfur-oxidizing symbionts	299.5	5.712	Hexamer
Dechloromonas aromatica	349.2	6.561	Hexamer
Ectothiorhodospira mobilis	311.7	5.491	Hexamer
Roseospirillum parvum	337.5	6.313	Hexamer
Gallionella sp. GWS1B	314.3	5.45	Hexamer
Insolitispirillum peregrinum	101	2.854	Dimer
Rhodobacteraceae_10_405	274	3.273	Hexamer
Rhodospirillum rubrum	100.8	3.352	Dimer
Rhodospirillaceae bacterium BRHc57	368.3	5.703	Hexamer
Sulfurivirga caldicuralii	218.3	4.896	Tetramer

Table 7-4. Statistics for data collection and refinement of *Sc* and BRH_c57 rubisco.

	Sc (L$_4$)	BRH_c57 (L$_6$)
Data collection		
Wavelength (Å)	1.000	0.97936
Resolution range (Å)	42.94 – 1.73 (1.79 – 1.73)	29.52 – 1.96 (2.03 – 1.96)
Detector Distance (mm)	170	250
Φ (deg.) collected / ΔΦ (deg.)	180/0.25	180/0.2
Exposure time (seconds)	0.25	0.2
Temperature of collect (Kelvin)	100	100
Data statistics		
Space group	P6$_2$22	P2$_1$
Unit-Cell parameters (Å)	a=b=133.05 and c=112.45	a = 74.80, b = 104.97 and c = 369.61. ß=93.03
Total reflections	122934 (12088)	776402 (68257)
Unique reflections	61467 (6044)	399609 (35839)
Multiplicity	19.4 (18.1)	1.9 (1.9)
Data completeness (%)	100 (99.9)	98.22 (88.1)
I/σ(I)	15.9 (0.7)	4.81 (1.26)
R$_{merge}$ (%)	0.123(1.088)	0.107 (0.518)
CC1/2	0.999 (0.353)	0.982 (0.41)
Structure Refinement		
Reflections used in refinement	54990 (4261)	399362 (35647)
Reflections used for R$_{free}$	1776 (127)	2001 (171)
R$_{factor}$ (%)	16.6 (37.1)	19.5 (28.6)
R$_{free}$ (%)	18.4 (38.3)	23.4 (32.7)
RMS from ideal geometry		
Bond lengths (Å)	0.004	0.009
Bond angles (º)	0.728	0.635
Average B-factor	40.92	35.02
Macromolecules	40.79	34.86
Solvent	42.41	36.56
Ramachandran Plot		
Favored region (%)	96.2	96.7
Outliers region (%)	0.2	0.2
PDB ID	7T1C	7T1J

7.3.5   Materials and Methods

7.3.5.1   Phylogenetic analyses

Form II and II/III amino acid sequences were originally compiled from Uniprot KB (https://www.uniprot.org/) using the search functions "rubisco" under protein name, and "cbbM" under gene name. The query results were assessed for inclusion based on sequence length and annotated oligomeric state. Form II/III sequences were included based on high sequence homology (>70%) to *M. burtonii* Rubisco. The resulting Uniprot KB sequence library was combined with the amino acid sequence library studied in Davidi et al. (2020)[348]. Rubisco sequences were then dereplicated at 97% amino acid identity using CD-Hit [366].

Sequences from the final library were aligned with MAFFT using default parameters (https://mafft.cbrc.jp/alignment/server/)[367]. Columns with >90% gaps were removed using TrimAI (http://phylemon2.bioinfo.cipf.es/). The evolutionary model most appropriate for constructing a phylogenetic tree was determined using Prottest 3.0[368]. A maximum-likelihood phylogenetic tree was constructed using RAxML-HPC BlackBox (v. 8.2.12) as implemented on cipres.org (default parameters with WAG model) with Form II/III sequences as the outgroup. The BOOSTER method was subsequently used to calculate the bootstrap branch support for the resulting phylogenetic tree (https://booster.pasteur.fr/) using "RAxML_bestTree" as the input reference tree and "RAxML_bootstrap" as the input bootstrap tree. All files used to create the phylogenetic trees are included on figshare.

7.3.5.2   Ancestral sequence reconstruction

Ancestral sequence reconstruction was performed with FastML v3.1 (http://fastml.tau.ac.il/) using the rubisco MSA and associated RAxML phylogenetic tree. Default parameters were selected, including branch length optimization, use of gamma distribution, indel reconstruction, and joint reconstruction computation. The sequences of the marginal reconstruction (including ancestral reconstruction of indels) were initially inferred using an indel cutoff of 0.2, 0.4, 0.6, 0.8, and 1.0. Amino acid sequence motifs and gaps from the Form II clade were most similar to the ancestral sequences constructed with either an indel cutoff value of 0.6 or 0.8, both of which produced near identical results. An indel cutoff of 0.6 was chosen for the final ancestral sequence reconstruction. All files used to create inferred ancestral sequences are included on figshare.

7.3.5.3   Relative amino acid evolutionary rate analysis

The relative evolutionary rates of amino acid residues found in hexameric form II rubisco were computed with Rate4Site v2.01 (https://www.tau.ac.il/~itaymay/cp/rate4site.html)[369]. First, the amino acid sequences for hexameric form II rubisco (including those identified in this study) were aligned with MAFFT using default parameters (https://mafft.cbrc.jp/alignment/server/)[367]. A maximum-likelihood phylogenetic tree was subsequently constructed using RAxML-HPC BlackBox (v. 8.2.12) as implemented on cipres.org (default parameters with WAG model). The MSA and associated phylogenetic tree were then used as input for Rate4Site to calculate the

relative conservation score for each site in the MSA.

7.3.5.4   Expression and purification of rubisco

Heterologously-expressed rubisco were purified in a manner similar to previously described methods[344,348]. BL21 DE3 Star competent *E. coli* cells (Macrolab, Berkeley, USA) were transformed with a pET28 plasmid containing the corresponding His14-bdSUMO-tagged rubisco sequence. Cells were grown at 37°C to OD_{600} ~0.6-0.8, followed by induction with 1 mM IPTG and further incubation overnight at 16°C. Cell cultures were then pelleted, resuspended in pH 8.0 lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM $MgCl_2$), and subjected to a freeze-thaw cycle. Thawed cells were then lysed using an Emulsiflex C3 (AVESTIN Inc., Ottawa, Canada). Lysate was clarified by centrifugation at 15,000xG, and soluble fractions were 0.44 μm-filtered before application to pre-equilibrated Ni-NTA resin for batch binding. Columns were washed twice, first with a 25 mM imidazole wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, 10% glycerol), followed by a 50 mM imidazole wash buffer (20 mM sodium phosphate, 300 mM NaCl, 50 mM imidazole, 10% glycerol). The column was then resuspended in pH 8.0 SUMOlase buffer (20 mM HEPES-OH, 100 mM NaCl, 1 mM DTT, 15 mM imidazole, 20 mM $MgCl_2$), and purified bdSENP1 was added and incubated overnight to facilitate tag cleavage[344,370]. Flow-through from the cleavage reaction was collected and analyzed by SDS-PAGE for purity.

7.3.5.5   Size-Exclusion Chromatography (SEC) coupled Small-angle X-ray Scattering (SAXS) with in-line Multi-Angle Light scattering (MALS) experiments (SEC-SAXS-MALS).

Rubisco was purified as described above and concentrated to 2-5 mg/mL. Concentrated rubisco was then activated with an excess of $NaHCO_3$ before sample analysis. SEC-SAXS-MALS data were collected at the ALS beamline 12.3.1 at Lawrence Berkeley National Lab (Berkeley, CA, USA; see Chapter 2)[10]. The X-ray wavelength was set at λ=1.24 Å and the sample-to-detector distance was 2075 mm resulting in scattering vectors (q) ranging from 0.01 Å^{-1} to 0.46 Å^{-1}. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. Data was collected using a Pilatus 3X 2M Detector (Dectris, Baden, Switzerland). Normalization and integration of each image was processed as previously described[9]. SEC was performed using a 1290 Infinity HPLC system (Agilent, Santa Clara, CA) coupled to a Shodex KW-803 column (Showa Denko, Tokyo, Japan). The column was equilibrated with a running buffer (20 mM HEPES-OH (pH 8.0), 300 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$) at a flow rate of 0.65 mL/min. 90-100 μL of sample was separated by SEC and the elution was monitored at 280 and 260 nm by an in-line Variable Wavelength Detector (VWD) (Agilent, Santa Clara, CA). MALS experiments were performed using an in-line 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab differential Refractive Index (dRI) detector (Wyatt Technology, Goleta, CA). System normalization and calibration was performed with bovine serum albumin using a 50 μL sample at 7 mg/mL in the same running buffer. The light scattering experiments were used to determine Molecular Weight (MW) across the principal peaks in the SEC analysis (**Figure S9**). UV, MALS, and dRI data were analyzed using Wyatt Astra 7 software to monitor the homogeneity of the sample across the elution peak complementary to the SEC-SAXS signal

validation. A purpose-built SAXS flow cell (see Section 2.3.2) was connected in-line immediately following the complementary spectroscopic techniques and two second X-ray exposures were collected continuously over the 25 min elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by radius of gyration (Rg) derived by the Guinier approximation, $I(q) = I(0) \exp(-q^2Rg2/3)$ with the limits qRg<1.5. The elution peak was mapped by comparing integral ratios to background and Rg relative to the recorded frame using the program SCÅTTER[54]. Uniform Rg values across an elution peak represent a homogenous assembly and were merged to reduce noise in the curve. Final merged SAXS profiles (Figure 7-10, Figure 7-11, Figure 7-13 and Figure 7-14), were used for further analysis including the Guinier plot which determined aggregation free state. The experimental SAXS profiles were then compared to theoretical scattering curves generated from atomistic models of *R. rubrum* (PDB: 5RUB) (Figure 7-13), the *Sc* tetramer (Figure 7-13), hexameric and dimeric *Gallionella* sp. states (Figure 7-10C, Figure 7-13), and engineered *Ip* enzyme (Figure 7-11C) using FoXS[371,372].

7.3.5.6   Crystallization and structural determination of rubisco

Ni-NTA-purified rubisco were further subject to anion exchange chromatography on a MonoQ 10/100 GL column and eluted by a linear NaCl gradient from 5 mM to 1 M. Fractions were analyzed by SDS-PAGE, followed by concentration and size exclusion chromatography on a Superose 6 Increase 10/300 GL, in a final buffer containing 100 mM HEPES pH 8, 100 mM NaCl, 25 mM $MgCl_2$, 5 mM $NaHCO_3$, and 1 mM DTT. Samples were activated as previously described before incubation with a tenfold molar excess of previously synthesized 2CABP[373].

The tetrameric *Sc* rubisco and the hexameric BRH_c57 rubisco were screened against the following crystallization screens: MCSG-1 (Anatrace), Crystal Screen, SaltRx, PEG/Ion, Index and PEGRx (Hampton Research) and Berkeley Screen[374]. Crystals of the *Sc* rubisco were found in 0.05 M citric acid, 0.05 M Bis-TRIS propane pH 5.0 and 16 % PEG 3,350. Crystals of the BRH_c57 rubisco were found in 0.2 M Magnesium formate pH 5.9 and 20 % PEG 3,350. Crystals from both enzymes were then placed in a reservoir solution containing 20% (v/v) glycerol and flash-cooled in liquid nitrogen.

The X-ray data set for the *Sc* rubisco was collected at the Berkeley Center for Structural Biology beamline 5.0.2 at the Advanced Light Source at Lawrence Berkeley National Laboratory, and the BRH_c57 data set was collected at beamline FMX at the National Synchrotron Light Source II at Brookhaven National Laboratory. The diffraction data were processed using the program Xia2[375]. The crystal structures of *Sc* and BRH_c57 were solved using molecular replacement with the program PHASER[376]. The atomic positions obtained from the molecular replacement were used to initiate model building using phenix.autobuild within the Phenix suite[377,378]. Structure refinement was performed using the phenix.refine program[379]. Manual rebuilding was done using COOT[380]. Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated with Phenix[378]. The stereochemical quality of the final models of *Sc* and BRH_c57 were assessed by the program MOLPROBITY[381]. A summary of crystal parameters, data collection, and refinement statistics can be found in Table 7-4. Structures and coordinates for *Sc* and BRH_c57 rubisco can be found in the PDB under accession IDs 7T1C and 7T1J, respectively.

7.3.5.7   Rubisco activity assays

Purified rubisco was used to determine catalytic properties as described previously[382], with some alterations to protein desalting and activation: concentrated protein aliquots were first diluted with activation mix containing 100 mM Bicine-NaOH pH 8.0, 20 mM $MgCl_2$, 10 mM $NaHCO_3$, and 1 % (v/v) Plant Protease Inhibitor cocktail (Sigma-Aldrich, UK). Rubisco was then activated on ice for 20 min before being used in $^{14}CO_2$ consumption assays at 25°C with CO_2 concentrations of 50, 100, 200, 300, and 400 µM. To determine K_O these CO_2 concentrations were combined with concentrations of either 0, 21, 40, or 70 % (v/v) O_2. k_{cat}^O was calculated from measured parameters using the equation $S_{C/O} = (V_C/K_C)/(V_O/K_O)$. k_{cat}^C was determined using measurements with 0% O_2. An aliquot of the activated protein was used for determination of Rubisco active sites via ^{14}C-CABP binding using the method of Sharwood et al[383]. Rubisco specificity was determined using the method of Parry et al[384]. Measurements using *T. aestivum* (bread wheat) rubisco were used for normalization as previously described, with a pKa of 6.11 used for calculations.

7.3.5.8   Protein Contacts Atlas analyses

Interface residues of the *S. caldicuralii* tetramer (PDB: 7T1C) and the *Gallionella* sp. Hexamer (PDB: 5C2C) were identified using Protein Contacts Atlas[351].

7.3.5.9   Site-directed mutagenesis experiments

Mutant rubisco was expressed and purified as previously described. Site-directed mutagenesis was conducted using an Agilent QuikChange Multi kit utilizing primers designed by the Agilent QuikChange Primer Design tool:

(https://www.agilent.com/store/primerDesignProgram.jsp).

7.3.5.10   Homology modeling

Rosetta CM was used to prepare a homology model for the input structure of the dimeric *Ip* enzyme[385]. MUSCLE was used for global sequence alignment during homology modeling[386]. Expanded sampling on sidechain chi angles resolved dimer-dimer interfacial interaction more accurately by using level 4 Rosetta rotamer libraries[387]. The flags and xml script used in homology modeling are available in supplementary information.

7.3.5.11   Symmetry definition

The symmetry definition was produced from make_symmdef_file.pl in Rosetta using the *Rhodospirillaceae* BRH_c57 structure as the input.

perl make_symmdef_file.pl -m NCS -p _49.pdb -a A -i C B -r 12 > _49.symm

7.3.5.12 Mutant selection

Mutation sites were identified by locating interfacial residues where BRH_c57 and *Ip* differ in protein sequence. Residues were defined as interfacial if 1) they were within 5.5Å of the opposite dimeric subunit, or 2) the sidechain points to the opposite dimeric subunit within 9Å. The mutation sites were manually screened, and seven sites were picked. All 128 combinations, each identified as a mutant, were modeled in silico as described below.

7.3.5.13 In silico mutation

In silico mutagenesis was performed on all 128 mutants. Monomeric rubisco structure was first extracted from the *Ip* homology model, and then applied with hexameric symmetry from the BRH_c57 structure (PDB: 7T1J). For each mutant, the residue(s) was mutated and the surroundings within a 12Å sphere of any mutation site were relaxed using the FastRelax protocol in Rosetta with level 4 rotamer libraries[387-390]. For each mutant, the structure was independently sampled 50 times and then ranked by its total energy (total score). The five samples with the lowest total energy were assessed with the number of dimer-dimer hydrogen bonds made, defined by a distance cutoff of 3.6Å. Upon manual inspection, mutants with the most hydrogen bonds were picked for experimental verification.

7.3.5.14 Other software

Multiple sequence alignments were generated using MAFFT and visualized with ESPript 3.0 [367,391]. Phylogenetic trees were visualized using Interactive Tree of Life v5[392]. UCSF ChimeraX was used for visualization of protein models and preparation of manuscript figures[393,394].

7.4 Case 3: Rigid monoclonal antibodies improve detection of SARS-CoV-2 nucleocapsid protein[§§]

7.4.1 Introduction

SARS-CoV-2 nucleocapsid proteins (NP) are critical for incorporating and packaging viral genomic RNA into mature virions. In infected cells, NPs are produced in large amounts from subgenomic mRNA and are present at the replication-transcription complexes (RTCs), the sites of RNA synthesis. The NP gene is relatively conserved, with a sequence identity of 91% and 50% to SARS-CoV and MERS-CoV, respectively, and is rather stable, as it acquires few mutations over time. [395,396] Although the NP from SARS-CoV-2 is abundant and highly immunogenic, [397-399] most SARS-CoV-2 detection assays use different spike protein regions as the antigen in immunoassays. This is mainly because antibodies against the spike protein are believed to be less cross-reactive [400] and are expected to correlate better with neutralizing capacity. [401] Testing for serum antibodies against NP from SARS-CoV-2 was suggested to increase diagnostic capacity. [398,402,403] However, serological assays cannot achieve diagnosis early in the onset of an infection because seroconversion occurs after 7-10 days in patients. [397,398,404]

Direct detection of viral proteins, often referred to as antigen-based detection, is more sensitive than serology assays in the case of SARS-CoV. [405] Antigen-based detection is amenable to use in rapid point-of-care lateral flow assays (LFA), which is another advantage. Thus far, antigen-based LFAs are significantly less sensitive than gold-standard RT-PCR but may approach RT-PCR's clinical sensitivity with further research and development. The choice of antigen, mAbs, and LFA protocols remain to be fully optimized for SARS-CoV-2.

The abundance and structure of NP in each virion provide a detection advantage over other antigen targets. NP is a 422 amino acid, 46 kDa phosphoprotein composed of two domains linked via a Ser/Arg rich linker with a short C-terminal region. NP dimerizes through its C-terminal domain (CTD). [406] The N-terminal domain (NP^{NTD}) is exposed and interacts with RNA. The independent NP^{NTD} and CTD domains do not have stable tertiary contacts in the absence of RNA. [406,407] In the presence of RNA, NP^{NTD} and CTD form a single bipartite RNA interaction site, which constitutes the basic building block of the nucleocapsid of SARS-CoV-2. [408,409] Abundance, stability, [406] and location at the surface of higher-order ribonucleoprotein assembly on the RNA [409,410] make the NP^{NTD} a viable antigen for the selection of highly specific mAbs for functional assays. NP is one of the early diagnostic markers in SARS-CoV-2, [411] and it has been detected one day before the onset of clinical symptoms in SARS infections. [412] Diagnostic fluorescence LFA immunoassays have been developed to detect SARS-Cov-2 NP protein in nasopharyngeal and nasal swab specimens. [413,414]

LFA protocols could take advantage of agglutination, a process in which antibodies mediate antigen-dependent aggregation into large particles. [415] The nature of the particles is influenced by antigen valency, enhancing antigen-antibody complex formation. [416,417] Agglutination is also a

---

factor when pairs of mAbs are used. LFAs that rely on a pair of mAbs that interact with different epitopes on an antigen have improved LFA sensitivity and specificity. [418] MAb-NP agglutination can serve to enhance the antigen-based detection limits against NP.

IgG flexibility, its importance in improving mAb recognition, and its influence on agglutination have remained uncharacterized. Although there have been several attempts by cryo-electron tomography [419-422] and negative stain (NS) electron tomography, [423] large scale flexibility measurements are often not amenable to single-particle techniques. In contrast, the resolution of small-angle X-ray scattering (SAXS) is sufficient, especially when atomic structures of individual components are available, to determine the conformational variability of the antigen-binding fragments (Fabs) in various antibodies,[424] including complexes with antigens or Fc-gamma receptors (FcγRs). [425,426] A previous study showed that the Fabs' conformational flexibility is derived from the inherent plasticity of the Fc-hinge regions in solution. [427] Rigidity of the hinges inversely correlates with, and can modulate, mAb agonistic potency, [428,429] and this highlights the importance of newer strategies to modulate antibody-agglutination. [430]

Here, we used SAXS and other biophysical techniques to structurally characterize mAbs that specifically bind the minimal NP^{NTD} region from a pool of 9 commercial mAbs raised against full-length NP. We correlated the observed flexibilities with super-structures formed when mAb pairs bind NP^{NTD}. Our structural insights have general implications for all antigen-antibody interactions. Simultaneously, a novel enzyme-linked immunosorbent assay (ELISA) protocol described here is intended to expedite the development of sensitive and selective antigen detecting LFAs, which could be applied in early diagnosis and epidemiological studies of SARS-CoV-2.

7.4.2   Results

7.4.2.1   mAbs against nucleocapsid N-terminal domain (NP^{NTD})

We used an integrative approach by size exclusion chromatography (SEC) coupled with SAXS and multi-angle light scattering (SEC-MALS-SAXS) to find mAbs that selectively bind minimal NP^{NTD}. SEC-MALS-SAXS experiments show that, from the pool of nine commercial mAbs raised against full-length NP, four antibodies (mAb1, mAb2, mAb4, and mAb8) bind NP^{NTD}. The SEC signal shifts with an increase in molecular weight (MW) (Figure 7-22A, Table 7-5), which shows that mAb1, mAb2, mAb4, and mAb8 form complexes with the NP^{NTD} in a 1:2 molar ratio. Additionally, the radius of gyration (Rg) values distinguish binder from non-binders (Figure 7-22B, Table 7-5). Final merged SAXS profiles for the corresponding SEC peak (Figure 7-27) were used to calculate pair-distribution functions (P(r)).

MAb binding of antigen is clearly distinguished by broad P(r) functions relative to those that remain unbound. The P(r) shape further provides information on the overall arrangement of mAb-antigen complexes (Figure 7-22B), which can be linked to the Fab's flexibility (Figure 7-23A). The first peaks in the P(r) function at r ~40Å arise from the approximate repeated distances across the Fc or Fab regions' length and breadth. The P(r) shoulder at r~80Å reflects the inter-domain distances between the Fc and Fab regions. Simultaneously, the divide between P(r) peak and shoulder reflects the Fabs' distancing, which correlates with the extended conformers' occupancy in solution. [425] The P(r) features and experimental Rg values (Figure

7-22B) allowed us to rank the inherent flexibility of mAbs, with mAb2 adopting the least and mAb4 the most extended states.

Table 7-5. SAXS, MALS, and SPR experimental parameters.

	mAbs (#)	Rg (Å)	Dmax	MW MALS/SAXS (kDa)	K_D mAb/+HRP (pM)	Simple SAXS ID
mAb + NP^{NTD}	1	52.9±0.5	165	193/177	1.3/11	BTQP75
	2	52.5±0.5	160	190/173	190	CBXGJF
	3	47.8±0.4	145	152/145		NEXZ6C
	4	54.1±0.5	180	184/170	11/28	ZMPE5M
	5	47.7±0.6	145	162/141		AMTYK1
	6	46.8±0.4	145	154/144		ULD5ED
	7	47.5±0.4	145	152/143		WHXQRV
	8	52.0±0.4	175	184/160		MSVSMP
	9	47.7±0.4	150	170/144		TQNW5I
free mAb	1	46.7±0.3	140	158/157		AKDRGZ
	2	46.9±0.3	145	150/147		W9GJYN
	4	50.0±0.3	155	150/167		PRDTAA
No-pair	1-4 +NP^{NTD}	53.2±0.5	180	190/169		ZGHQLG
Pair	1-2 +NP^{NTD}	73.9±0.9	~300	390/397		WNHK6M
	2-4 +NP^{NTD}	67.6±0.6	~280	370/320		UJ5ICU
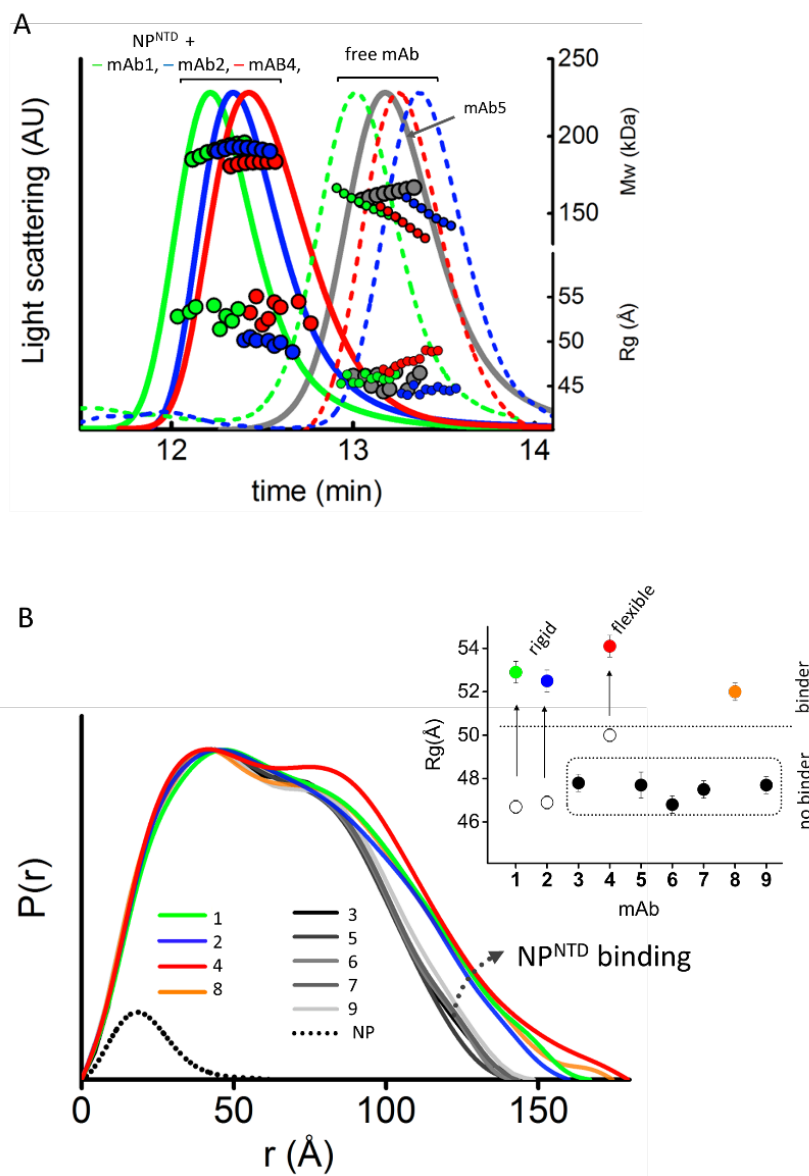
Figure 7-22. SEC-MALS-SAXS identifies mAbs that bind to NPNTD. (a) SEC-MALS-SAXS chromatograms for free and NPNTD bind mAb1, 2, and 4 (green, blue and red lines). Chromatogram for mAb5 + NPNTD (gray) sample is included for comparison to a no-binder. Solid lines represent the light scattering signal in arbitrary units, while symbols represent molecular mass (top) calculated from MALS and Rg values (bottom) for each collected SAXS frame versus elution time. (b) P(r) functions calculated for the experimental SAXS curves for all tested mAb + NPNTD samples (colored as indicated). The P(r) functions are normalized at the maximum. The experimental P(r) function for NPNTD alone is shown for the comparison and normalized relative to the MW estimated by SAXS.37 Inset: Experimental Rg values determined by Guinier plot for the experimental SAXS curves of mAb +NPNTD mixtures (solid dots) and mAb1, 2, and 4 (circles) indicate binder and no binder. Experimental SAXS curves for mAbs + NPNTD and free mAb1, 2, and 4 are shown in Figure 7-27 and Figure 7-23B, respectively.
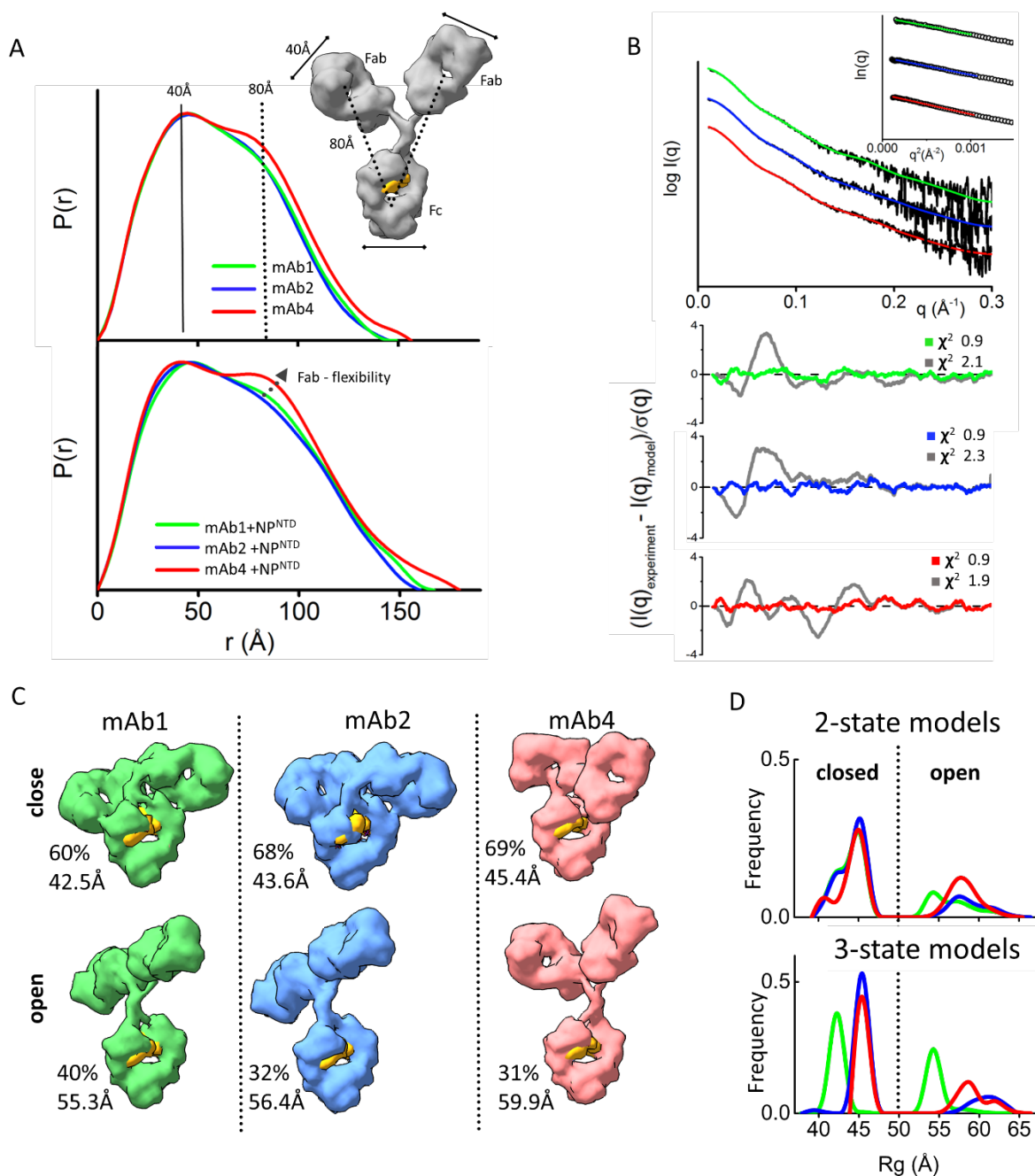
Figure 7-23. The flexibility of the NPNTD-binding mAbs. (a) P(r) functions for free mAb 1, 2, and 4 (top) and their complexes with the NPNTD normalized onto their maxima. The P(r) shoulder at r ~ 80 Å indicates the Fab-Fc separation described within the atomic model of IgG1 (inset). P(r) peak at 40 Å corresponds to the average size across Fc or Fab regions. (b) Experimental SAXS profiles of free mAbs 1, 2, and 4 (black) and theoretical SAXS profiles calculated from their respective two-state atomistic models (green, blue, and red) are shown in the panel. Residuals (Experiment/Model) for the fits of two-state models (green, blue, and red) are shown together with the best single model (gray) and indicate that the two-state model is

required to match the experimental SAXS curves. (c) Two-state models for free mAb 1, 2, and 4 are shown together with the corresponding weights in % and Rg values. The Rg values and weights of mAb4 further confirm a larger separation between the Fc and Fab region. The atomistic models are shown as molecular envelopes at 10 Å resolution. The glycan-moiety in the Fc region is colored yellow. (d) The Rg distributions of the top 300 selected multistate models are shown for two-state (top panel) and three-state models (bottom panel) of free mAb 1, 2, and 4 (green, blue, and red).

7.4.2.2   mAbs with distinct flexibility of the Fab domains

Interpretation of SAXS and P(r) functions is further enhanced by available atomic models of mAbs. While the crystal structure of intact human IgG1 antibody (PDBID:1HZH) does not fit the SAXS data, it forms the basis for creating an ensemble of conformations. We used the program BILBOMD [431] to explore the Fab regions' conformational space relative to the Fc. BILBOMD performs minimal molecular dynamics (MD) simulations on the Fc-hinge regions at very high temperature, where the additional kinetic energy prevents the Fabs from becoming trapped in a local minimum. This conformational sampling provides a pool of atomistic models (> 10,000) from which SAXS curves are calculated [291] and compared to the experimental curve. MultiFoXS algorithm [284] is used to identify the weighting of multistate models that fit the experimental data.

At least two distinct conformations are required to fit the SAXS data measured for the three mAbs that bind antigen (mAb 1, 2, and 4). A single conformation from BILBOMD failed to adequately match our measured SAXS profiles with poor goodness of fit (χ^2 2.1, 2.3 and 1.9) (Figure 7-23B). For each mAb, we found significant improvement in the SAXS fit by selections of two-state models with similar goodness of fit for all three mAb (χ^2 0.9) (Figure 7-23BC). At the same time, the three-state models do not improve SAXS fit (χ^2 0.9). To estimate the number of conformational states in solution, we examined the R_g distribution[282] for the top 300 selected multistate-models. The R_g distribution of the two-state models (Figure 7-23D) has two peaks: one corresponding to closed conformations at 43-46 Å and the other corresponding to open conformations at 53-65 Å. For three-state models (Figure 7-23D), the R_g distribution also has two peaks, suggesting that mAbs adopt two states in the solution, a closed conformation and flexible-open conformation. The area under the Rg-distribution curve indicates a higher population of mAb4-open conformers, whereas the shift in peaks suggests the superior rigidity of mAb1. For a better representation of the conformational space that the mAbs occupy, the top selected two-state model is shown in Figure 7-23C for each of the mAbs that bind antigen. MAb-binders (mAb 1, 2, and 4) show differences in conformational variability between two-states. Both mAb4 open and closed conformers show significant separation between the Fc and Fab regions (Figure 7-23C) relative to those found to fit data from the other two mAbs. This difference provides further insight into the prominent P(r) shoulder observed for mAb4 (Figure 7-23A).

The same feature, indicating additional mAb4 flexibility, is observed in the P(r) functions when NP^{NTD} is present (Figure 7-23A). A more distinct separation of the P(r) shoulder in the mAb4-NP^{NTD} complex and free state (Figure 7-23A bottom) indicates larger distancing of Fab from Fc. On the other hand, smaller P(r) shoulders (Figure 7-23A) together with reduced experimental Rg

values (Figure 7-22B) of the mAb1-NP^{NTD} and mAb2-NP^{NTD} complexes correlate with the P(r) shapes of free mAb1 and mAb2, which suggests rigidity of the antibodies. Comparable Fab-flexibility between free and NP^{NTD}-bound states agree with previous MD simulations showing only minor allosteric communication between Fab and Fc domains upon antigen binding. [432]

7.4.2.3   Fab flexibility correlates with a sandwich or linear pairing of mAbs

MAb pairs that simultaneously bind the same NP^{NTD} through different epitopes are also readily distinguished from pairs that compete for the same epitope by SEC-MALS-SAXS. Based on the SEC elution profile and MALS-determined MW across the SEC peak, we show that the NP^{NTD} does not bridge mAb1 and 4 (Figure 7-24A). Thus, mAb1 and 4 compete for binding to NP^{NTD}. In contrast, higher mass species were formed by mixing mAb2-NP^{NTD} with either mAb1 or mAb 4, showing that mAb1-2 or mAb2-4 are pairing through simultaneous binding with NP^{NTD} at different epitopes (Figure 7-24A). Control experiments show that neither mAb1-2 nor mAb2-4 mixtures form larger complexes in the absence of NP^{NTD} (Figure 7-28).

Each mAb pair binds NP^{NTD} in different stoichiometries and orientations. Mass by MALS and SAXS from the main elution peak show the complex formed by mAbs1-2 is ~390 kDa, while the mAbs2-4 is ~370 kDa, which corresponds to two antibodies bound by three or two NP^{NTD} molecules, respectively. Also, the orientation of binding between the pairs is very different. The Rg of mAb1-2-NP^{NTD} is 74 Å relative to the 68 Å measured for mAb2-4-NP^{NTD} (Figure 7-24A, bottom right axis). Furthermore, Rg changes are accompanied by a shift in the secondary peak in the P(r) distribution (100Å vs. 80Å). To gain insights into the structures these mAb pairs form, we reconstructed SAXS envelopes for both mAb1-2-NP^{NTD} and mAb2-4-NP^{NTD}. The envelopes for mAb2-4-NP^{NTD} show a sandwich-like assembly with a hollow feature in the center of the model, whereas the mAb1-2-NP^{NTD} adopts a linear arrangement.

We manually superimposed the SAXS envelopes with their corresponding mAb-atomistic models to approximate the overall arrangement of mAb-pairs. The sandwich-like arrangement of mAb2-4-NP^{NTD} matches the SAXS envelope and shows two antigens bound between two Fabs. The SAXS envelop of mAb1-2-NP^{NTD} matches a linear arrangement of the antibodies where only one NP^{NTD} is shared between mAb1-2 (Figure 7-24C). The shapes and models of the complex provide insights into the P(r) distributions' shifts.

We postulate that the difference in orientation fundamentally relies on differences in the flexibility of the mAbs. The mAb2-4 pair contains the flexible mAb4 and shows a closed and capped arrangement around two antigens. MAb4's flexibility allows the Fabs to stretch to accommodate two NP^{NTD} molecules' binding located on the Fabs of mAb2. In contrast, the more rigid mAb1-Fab regions restrict the Fabs' positioning onto the NP^{NTD} located on the mAb2. Thus, the relative rigidity of both mAb2 and mAb1 enforces the linear arrangement of the mAb1-2-NP^{NTD} complex.

The linear antibody-antigen arrangement of the mAb1-2-NP^{NTD} complex should permit further networking of multiple mAbs through the uncovered epitopes of the NP^{NTD} molecules bound to the outermost Fab regions. Indeed, there is a notable presence of very large complexes (~1 MDa) in the mAb1-2-NP^{NTD} sample (Figure 7-24A), suggesting further elongation of the complex by extending the rigid linear arrangement (Figure 7-24BC). The mAb2-4-NP^{NTD} also shows a small

amount of very large complexes. However, the low SAXS signal (Figure 7-24A) did not allow us to determine this species' overall arrangement. These observations suggest flexibility of mAbs is a factor in the agglutination of mAb - antigen complexes.
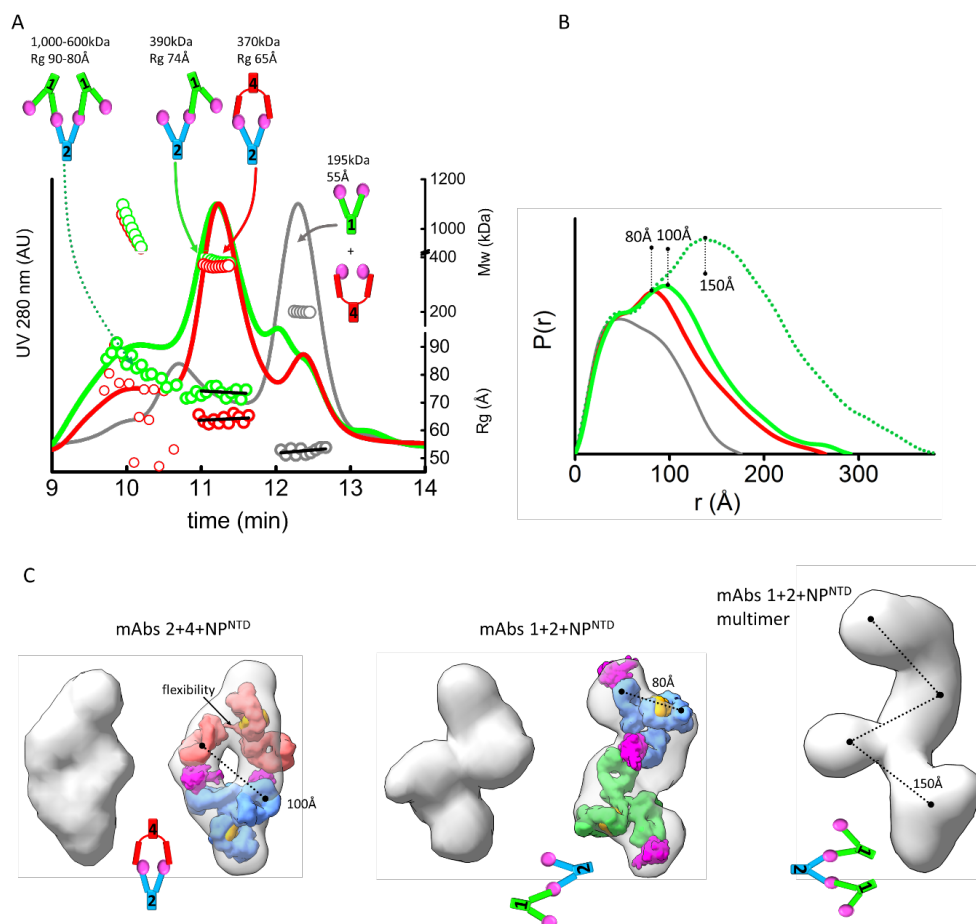
Figure 7-24. MAb linear or sandwich pairing depends on inherent flexibility. (a) SEC-MALS-SAXS chromatograms for the mAb1-2-NPNTD (green), mAb2-4-NPNTD (red) and mAb1-4-NPNTD (gray) samples. Solid lines represent the UV 280 nm signal in arbitrary units, while symbols represent molecular mass (top) calculated from MALS and Rg values (bottom) for each collected SAXS frame versus elution time. (b) P(r) functions calculated for the experimental SAXS curves for the main SEC peak of mAb12-NPNTD (green), mAb2-4-NPNTD (red), mAb1-4-NPNTD (gray), and early SEC shoulder of mAb1-2-NPNTD (green dots). The P(r) functions are normalized at the r = 40 Å. The P(r)-maxima peaks are indicated. Experimental SAXS and Guinier plots are shown in Figure 7-27. (c) Average SAXS envelopes obtained for mAb2-4-NPNTD, mAb1-2-NPNTD complexes were calculated using a P2 symmetry operator. Average SAXS envelopes calculated using a P1 symmetry operator are shown in Figure 7-29. A single representative envelope was manually superimposed with compact conformers of mAb1 (red), mAb2 (blue), and mAb4 (green) taken from the two-state model of free mAbs (see Figure 7-23C). The structure of NPNTD (magenta; PDB ID: 6VYO) was manually docked at the proximity of the CRD3 -Fab region. Additionally, the SAXS envelope obtained for the larger multimer of mAb1-2-NPNTD determined in P1 symmetry is shown.

210
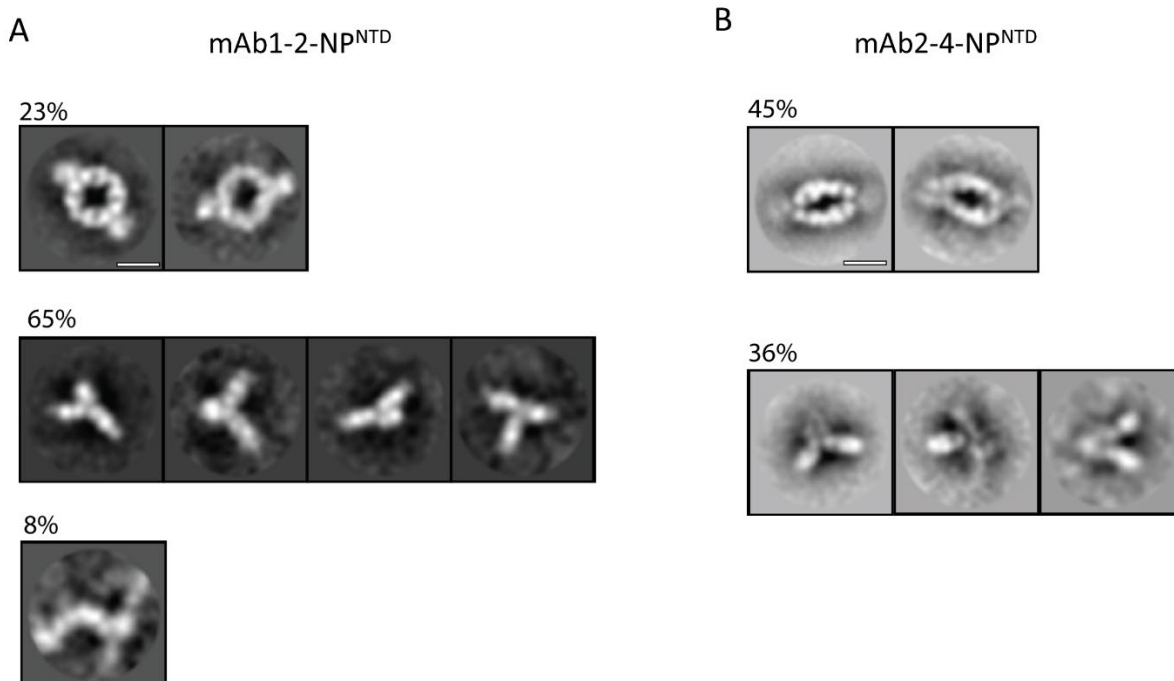
Figure 7-25. TEM 2D class averages highlight mAb-NPNTD pairing differences. Representative 2D class averages from NS-TEM data for mAb1-2-NPNTD (a), and mAb2-4-NPNTD (b). The scale bars are 10 nm.

7.4.2.4   Sandwich and linear mAb-pairing observed by TEM

To further support our solution scattering results, we applied freshly prepared samples of either mAb1-2-NP^{NTD} or mAb2-4-NP^{NTD} to transmission electron microscopy (TEM) grids for TEM analysis (Materials and Methods Section 7.4.5; Figure 7-25). While both samples show the mAb-NP^{NTD} pairs can form the sandwich arrangement, these are formed almost twice as readily in the more flexible mAb2-4-NP^{NTD} (45%, Figure 7-25B) compared to the more rigid mAb1-2-NP^{NTD} (23%, Figure 7-25A). Further, the flexible mAb2-4-NP^{NTD} sandwich appears elongated or more 'relaxed,' while the more rigid mAb1-2-NP^{NTD} sandwich appears circular and more 'strained.' There is a significant presence of single antibodies or single antibodies bound by NP^{NTD}, which is evident to a lesser degree in the solution scattering (Figure 7-24A – right complex peaks). However, the classification method, relying on the alignment of identical features, may overestimate the number of mAb monomers and underestimate the number of linear arrangements of the populations. A linear arrangement was also evident in the mAb1-2-NP^{NTD} mixture only, albeit at only 8%. While these complexes clearly highlight differences in mAb pairing that agree with the solution scattering (Figure 7-24), the samples were relatively dilute. Thus, we sought to image the complexes at a higher concentration with an extended incubation time (see Materials and Methods Section 7.4.5).

As shown in Figure 7-30, it is immediately apparent that the higher concentration, longer incubation period samples have a significantly higher prevalence of the elongated sandwich

arrangement for the flexible mAb2-4-NP^{NTD} pair (Figure 7-30B), at 28%, versus 6% for the more rigid mAb1-2-NP^{NTD} pair (Figure 7-30A). However, under these conditions there is little discernible difference in the percentage of extended linear arrangements (i.e., polymerized) in the mAb1-2-NP^{NTD} sample (Figure 7-30A) compared to the mAb2-4-NP^{NTD} sample (Figure 7-30B; 11% vs. 10%). Collectively, these TEM data support the solution scattering data, which demonstrate that a rigid antibody-antigen pairing has a higher propensity to form extended linear complexes, while a flexible antibody-antigen pairing facilitates the formation of sandwiched complexes.

7.4.2.5 SPR kinetic analysis revealed comparable picomolar affinities of all antibodies

To compare the relative affinity of each mAb for antigen, we performed binding kinetic assays. In addition, we performed assays on horseradish peroxidase (HRP)-conjugated mAbs in preparation for ELISAs, described below. Due to the high affinities of the mAbs, we opted to use a kinetic titration (single cycle kinetics) strategy and avoid problematic regeneration steps (Materials and Methods Section 7.4.5). We measured the binding kinetics of mAb 1, 1-HRP, 2, 4, and 4-HRP by surface plasmon resonance (SPR) (Figure 7-31). All antibodies (unconjugated and HRP-conjugated) had high-affinity constants (K_D) in the picomolar range (Table 7-6). The K_D of HRP-conjugated mAb1 and 4 are very similar, at 11 and 28 pM, respectively. The percentage activity of the HRP-conjugated antibodies is lower than unconjugated, suggesting that conjugating HRP on the antibodies affects the percentage of available antibodies for interaction on the SPR sensor chip. The possibility exists that this effect could also be present in the chip-free solution-based ELISA. However, the high concentration of HRP-conjugated antibodies used (0.4 mg/mL; Methods Section 7.4.5), relative to the picomolar affinities, represents a large excess of functional, high-affinity HRP-conjugated antibodies in the ELISA. Therefore, the antibodies have comparable kinetics, effectively excluding them as explanations for functional outcomes.

7.4.2.6 A modified ELISA protocol "boosts" the signal of the linear mAb arrangement

We sought to assess the consequences of the observed mAb linear arrangement vs. sandwich pairing (Figure 7-24) on detection limits. Since mAb2 pairs with mAb1 and mAb4, we used mAb2 as the NP^{NTD} capture antibody and conjugated HRP to mAb1 and mAb4 (1-HRP, 4-HRP) to serve as the detection antibodies. We hypothesized that the linear arrangement of mAb1-2-NP^{NTD} could facilitate a higher ratio of detection-to-capture mAbs (two or more 1-HRP antibodies to capture mAb2) on the plate, leading to a boost in the signal. This would contrast with the sandwich pairing of mAb4, which closes off the further binding and constrains assembly to a 1:1 ratio of 4-HRP to mAb2. To test this hypothesis, we developed a modified ELISA protocol.

To enhance detection, we modified the standard ELISA protocol. The two main differences between this and a standard ELISA are: 1) The detection HRP-conjugated mAbs are added directly on top of the samples during the incubation period that is typically used for the capture of the antigen only, and 2) Free (non-plate-bound) mAb2 is "spiked" into the detection HRP-conjugated mAb solutions before their addition on top of the samples. We rationalized that

adding antigen simultaneously with detection antibodies would initiate maximal polymerization and that a later "spiking" in of mAb2 would further extend "networking" of the linear mAb1-2-NP^{NTD} arrangement (Figure 7-24C – middle/right panels), whereas, the sandwich pairing of mAb2-4-NP^{NTD} does not allow the polymerization of antibodies (Figure 7-24C – left panel).
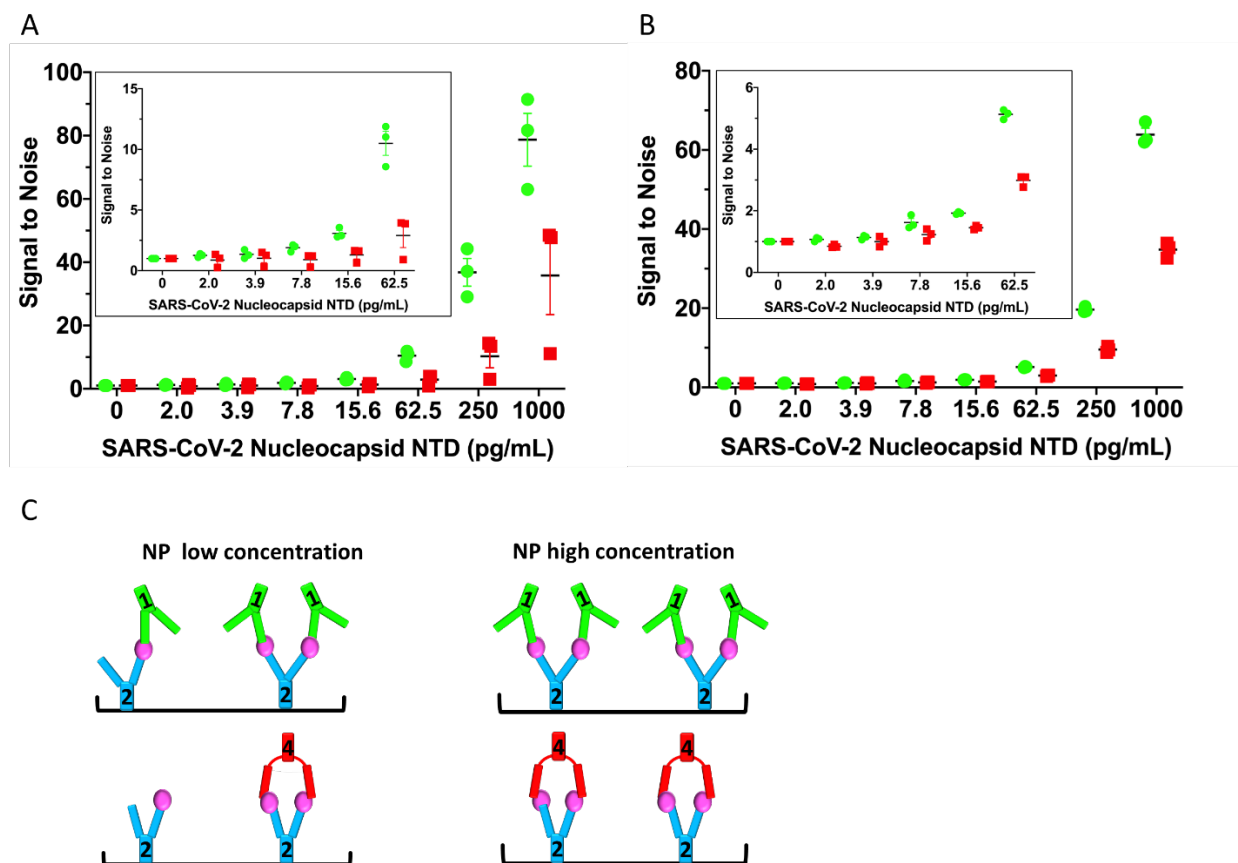
Figure 7-26. Linearly arranged mAbs show boosted signal in modified ELISA. (a) A modified ELISA where the detection HRP-conjugated mAbs (1-HRP in green, 4-HRP in red) are added directly on top of the samples during the NPNTD capture incubation period. Free (non-plate-bound) mAb2 is "spiked" into the detection HRP-conjugated mAb solutions before their addition on top of the samples. The corresponding standard control ELISA protocol run in parallel on the same plate is shown in Figure 7-32A. (b) Repeat of the experiment conducted in (a), with a corresponding control ELISA protocol run in parallel on the same plate with a longer mAb-HRP-sample incubation period, shown in Figure 7-32B. (c) Schematic of low versus high concentration of NPNTD in samples. In both experiments, the 1-HRP that forms the more rigid linear arrangement in the unconjugated form (mAb1-2-NPNTD) shows an ~2-fold increased ELISA signal, relative to 4-HRP, that forms a sandwich arrangement in the unconjugated form (mAb2-4-NPNTD). Each NPNTD concentration was run in triplicate, and the standard error of the mean for each is included.

Following this protocol, we observed improvements in detection limits using mAb1-2-NP^{NTD} relative to mAb2-4-NP^{NTD} (Figure 7-26). In repeated experiments (Figure 7-26AB), the signal of 1-HRP is ~2-fold higher than 4-HRP, although the magnitude of the effect is diminished with decreasing concentration of antigen (Figure 7-26AB - insert; Figure 7-26C). The two independent experiments (Figure 7-26AB) each had a control experiment done in parallel on the same plate (Figure 7-32A and B, respectively). The first control experiment follows the standard ELISA protocol, where the plate is washed prior to the addition of the mAb-HRP for a 20-minute incubation (Figure 7-32A). To control for the longer incubation time of the mAb-HRP with the samples in our modified ELISA protocol, we ran an additional control (Figure 7-32B), where the mAb-HRP had a longer incubation time of 1.5 hrs. No "boost" of the 1-HRP signal over the 4-HRP level was seen in either control experiment (Figure 7-32AB). Simultaneously, there was a general elevation of both signals in the control with the longer incubation time (Figure 7-32B). These results collectively demonstrate the ability to capitalize on the linear-mAb arrangement functionally, which results from the structural rigidity of the antibodies (Figure 7-22, Figure 7-23, and Figure 7-24).

Further, we were interested in whether the structurally enforced functional "boost" effect could be maintained in the presence of a virion-disrupting detergent (Figure 7-33) since NP is present inside virions. SARS-CoV-2 virions are not lysed adequately in the presence of 0.5% Tween-20, a detergent commonly used in ELISAs that is present in our protocol at a lower concentration, 0.05% (Methods Section 7.4.5), but are effectively lysed in the presence of 0.5% triton X-100. [433] Therefore, we used our same modified ELISA protocol that demonstrated the "boost" (Figure 7-26AB), except that we used phosphate-buffered saline (PBS) pH 7.4 plus 0.5% triton X-100 as the sample dilution buffer, instead of PBS pH 7.4 alone. The presence of triton X-100 reduced the "boost", although it is still detectable (Figure 7-33). Interestingly, the presence of triton X-100 appears to have increased the overall limit of detection (LOD) to lower than 0.4 pg/mL. In contrast, it is clearly not this low in the detergent's absence (compare 0 and 2 pg/mL in Figure 7-26AB with 0 and 1.7 pg/mL in Figure 7-33). Further improvements could be gained to maximize both the "boost" and detergent effects. Together, these results suggest that combining our modified ELISA protocol with the presence of a SARS-CoV-2 virion lysing concentration of triton X-100 leads to a highly sensitive ELISA, with great potential for further diagnostic development.

7.4.3   Discussion

The ongoing SARS-CoV-2 pandemic has highlighted the need for sensitive point-of-care diagnostics (POCs), which are primarily antibody-based technologies. [434] Currently, mAbs are widely used to detect antigen molecules, including the nucleocapsid protein from SARS-CoV-2. The US Food and Drug Administration recently authorized a lateral flow antigen test as the first over-the-counter, fully at-home diagnostic test for the qualitative detection of SARS-CoV-2 nucleocapsid antigens. [435] However, there is an urgent need to improve the detection limit of these diagnostic devices that use various types of colorimetric mAb-based assays. [434]

Multiple approaches, like florescent immunoassays, [436-438] nanoparticle luminescence, [439,440] or magnetic beads as the antibody support surface, [441] are used to enhance antigen detection. In immunoassays and RT-PCR, the detection is signaled through chemical conjugation to an

enzyme or nanoparticle that drives a colorimetric reaction, a fluorophore, or another moiety. However, a limitation of existing immunoassays in detecting antigens relative to RT-PCR is the lack of exponential amplification of signal when probes detect an antigen. Immunoassays mainly rely on antibody-antigen binding at a 1:1 ratio. An immunoassay diagnostic with a greater detection-to-capture antibody ratio will also have a greater signal-to-antigen ratio, effectively enhancing overall specific antigen detection.

Despite the widespread use of antibodies in diagnostics and treatments, an understanding of structural properties that affect antibody-antigen interactions in their aqueous environment is currently insufficient to guide optimization for these purposes. Here, we describe how antibodies' inherent flexibility leads to distinct antigen-binding arrangements that drive different functional outcomes in a SARS-CoV-2 detection ELISA. We show that we can rapidly assess new antibody-antigen interactions using SEC-MALS-SAXS to identify pairs that bind in a linear arrangement (Figure 7-24), which results in a more sensitive detection assay (Figure 7-26). A critical benefit of using the SEC-MALS-SAXS approach is its ease of use and the ability to study antibody-interactions in solution. It has previously been shown that other techniques that rely on grids (electron microscopy) or crystals often do not reflect the dynamic nature of antibodies in solution. [421] However, electron microscopy can be used to visualize the overall arrangement of larger mAb complexes. Indeed, our TEM experiment supports our solution-based models, demonstrating different pairing of mAbs with NP^{NTD}. Interestingly, while both TEM and SEC-MALS-SAXS clearly demonstrate differences in pairing, the linear arrangements were better detected in the latter, highlighting the importance of assessing protein-protein interactions in their naturally dynamic solution states. Thus, these experiments further affirm the SEC-MALS-SAXS technique as a means to visualize mAb dynamicity and networking in an efficient way.

A central finding in this work is the observation and rationalization of how mAb flexibility can affect larger assemblies of more than one mAb with or without antigen. Many mAbs are abandoned as formulations despite a high affinity for their antigen because of a propensity to aggregate. SEC-MALS-SAXS is a rapid method that allows separation of larger-scale assemblies from single mAbs and, therefore, an interrogation of aggregation propensity on the structural properties of the underlying single mAb. Further, SAXS provides the resolution to distinguish rigid and flexible mAbs in solution. Flexibility can arise not only from primary sequence differences, but it may also occur due to glycosylation. This may have been a factor in our studies, as we used both mouse (mAb1, mAb2) and rabbit (mAb4) host antibodies. [442] In light of our results on the impact of flexibility, further studies to assess the effects of glycosylation may be worthwhile, as glycosylation can be adjusted.

Having identified two mAbs (mAb1 and mAb4) with differing degrees of flexibility (Figure 7-23) that can both bind SARS-CoV-2 NP^{NTD} simultaneously with a third antibody mAb2, we were able to contrast the larger assemblies composed of mAb1-2 NP^{NTD} versus mAb2-4 NP^{NTD} (Figure 7-24). Further analysis confirmed the existence of two different binding modes of antibody-antigen-antibody: sandwich, and linear (Figure 7-24). The linear mode suggested further polymerization might be possible. This polymerization would be considered aggregation when interpreted by other methods. However, we sought to use this propensity to overcome the limitation on amplification inherent in antibody-based diagnostics.

Based on the above observations, we developed a modified ELISA. We used mAb2 as a common capture antibody and mAbs 1 and 4 as detection antibodies. We saw an increased signal for the mAb1-2 pairing relative to mAb2-4, where mAb1 and mAb4 were coupled to HRP, a common signal-generating enzyme used in the ELISAs. [443] To control for changes in binding kinetics, we used SPR to demonstrate that the near-equivalent binding affinities for nucleocapsid are maintained after HRP-conjugation of the two antibodies. The modified ELISA that used the linearly arranged mAb1-2 pair consistently generated a larger signal than the sandwich mAb2-4 pair. Further, the signal of 1-HRP is ~2-fold higher than 4-HRP, and the effect is diminished with decreasing concentration of antigen, which supports our hypothesis.

Having made gains in detection by considering the structural properties of mAbs, more optimization is likely possible. Introducing further rigidity in mAb2 through glycosylation modifications, binding factors like protein A or G, detergents, or other metabolites could enhance further networking and, therefore, detection. The positive signal line in LFAs is often generated by antibodies conjugated to colloidal gold or latex, which accumulate into pink or blue lines, respectively. [444] These are meant for visual inspection by non-experts in POC devices. A previous study demonstrated that the detection limit of an LFA could be lowered 3-fold, from 3.1 ng/mL to 0.9 ng/mL for detection of aflatoxin B2 in food, through noallenlab_SEC-SAXS091018 report n-covalently clustering (16 nm diameter) gold nanoparticles for a visual readout. [444] Thus, the clustering of signal molecules coupled to antigen-specific antibodies is a viable strategy for lowering the LOD in LFAs. Our study shows that this can be achieved without introducing an additional factor by taking advantage of the antibodies' structural rigidity. A survey of commercial ELISAs suggests a common LOD of 100 pg/mL, with the most sensitive being 0.01 pg/mL for protein analytes. [445] Many antibodies have been identified against NP and other antigen targets from SARS-CoV-2. SEC-MALS-SAXS could be applied to hundreds of mAbs in a short amount of time to identify the most rigid. By combining this novel strategy with other optimization methods (e.g., tuning antibody affinities, selecting signaling molecule/moiety), the LOD of standard LFA POC devices could achieve as yet unattained sensitivity for current and future pathogens.

7.4.4 Supplementary Material
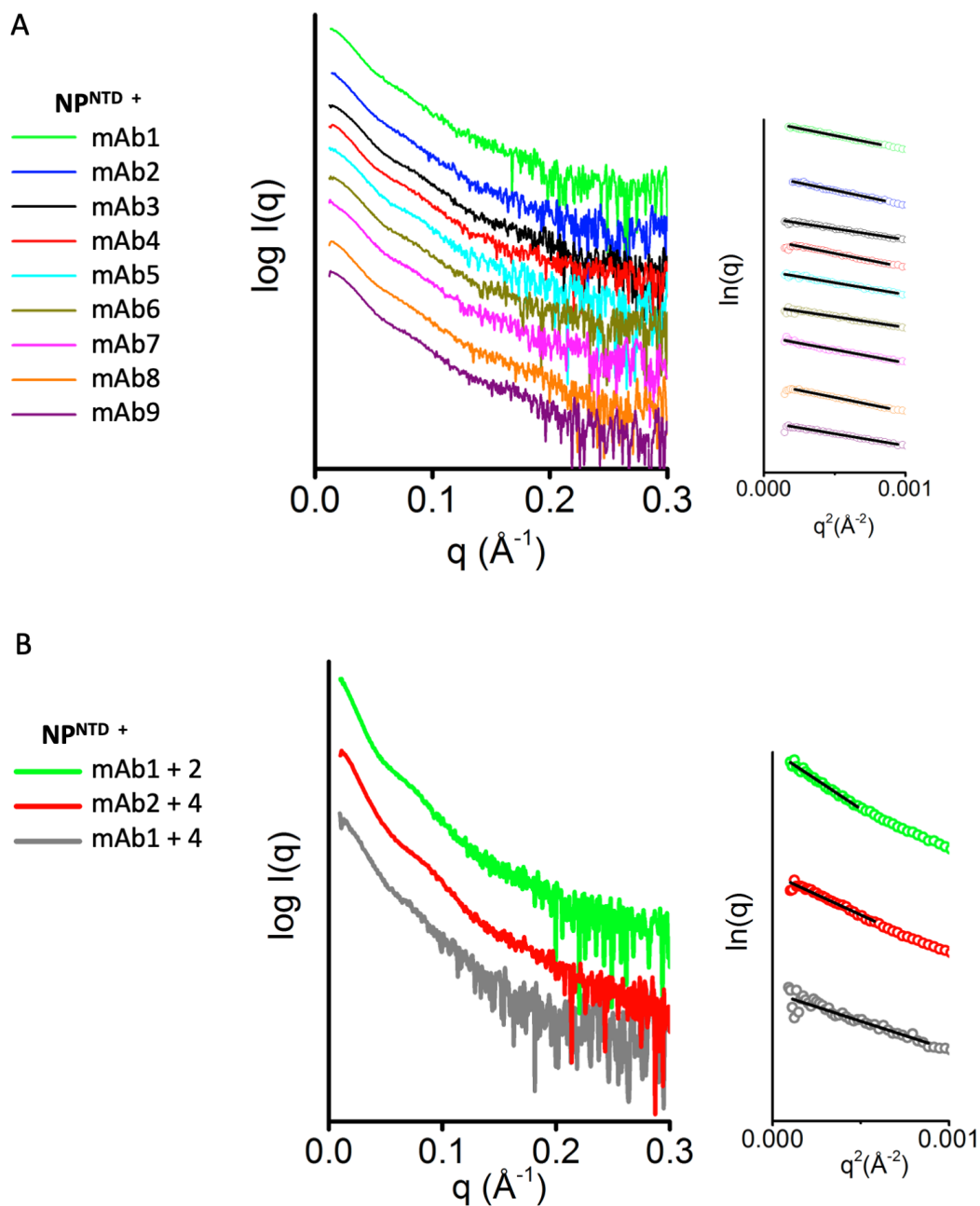
Figure 7-27. SAXS profiles. (A) Experimental SAXS curves of all tested mAb + NP^{NTD} samples (colored as indicated). Right panel: Corresponding Guinier plot with q*Rg <1.5 limit. (B) Experimental SAXS curves for the main SEC peak of mAb1-2-NP^{NTD} (green), mAb2-4-NP^{NTD} (red), mAb1-4-NP^{NTD} (gray). (A-B) Guinier plots were used to determine Rg values used in Figure 7-22B-inset and listed in Table 7-5.
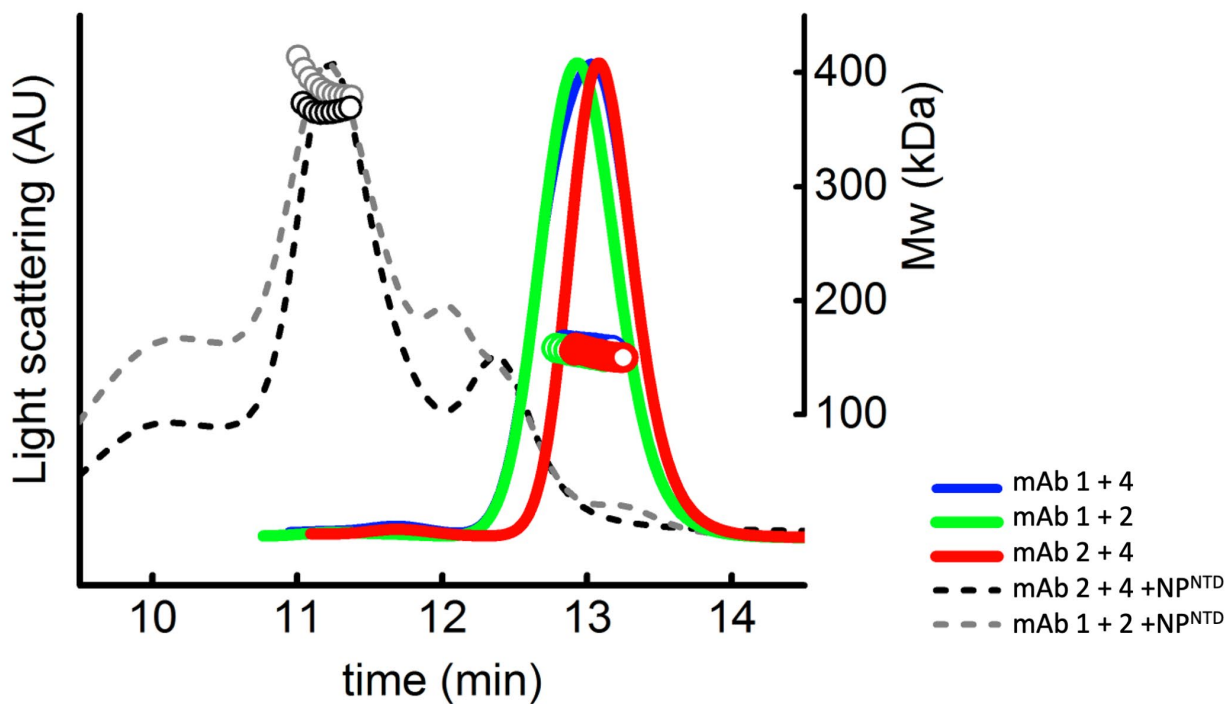
Figure 7-28. mAb pairs do not form a large assembly in the absence of NP^{NTD}. SEC elution profile (lines) and MALS-determined molecular weight (circles) across the SEC peak for mAb1+4, mAb1+2, mAb2+4 in absence of NP^{NTD} (in comparison to the mAb1+2, mAb2+4 in presence of NP^{NTD} (colored as indicated).

mAbs 2+4+NP^{NTD}

P2   P1

mAbs 2+1+NP^{NTD}

P2   P1
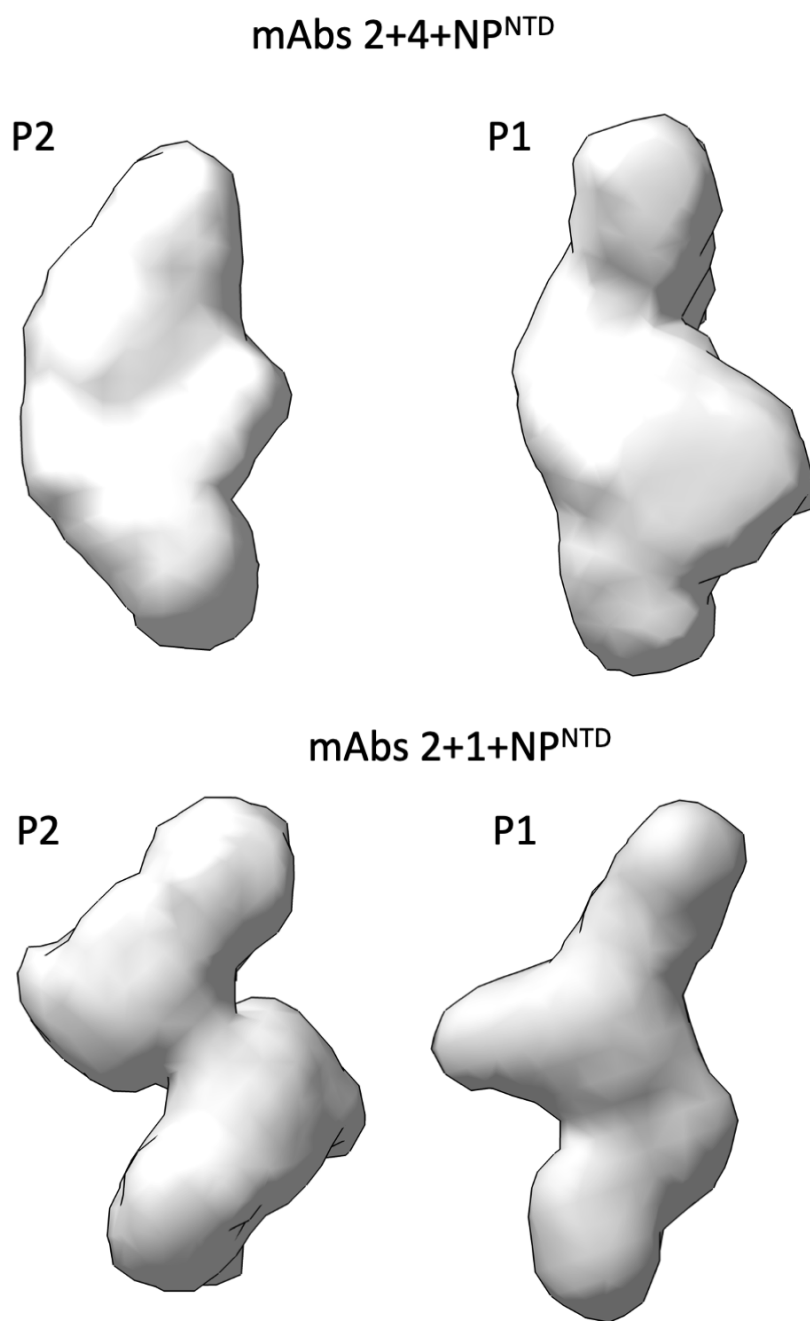
Figure 7-29. Average SAXS envelopes. Comparison of average SAXS envelopes obtained for mAb2-4-NP^{NTD}, and mAb1-2-NP^{NTD} complexes reconstructed using a P2 (left) and P1 (right) symmetry operator.
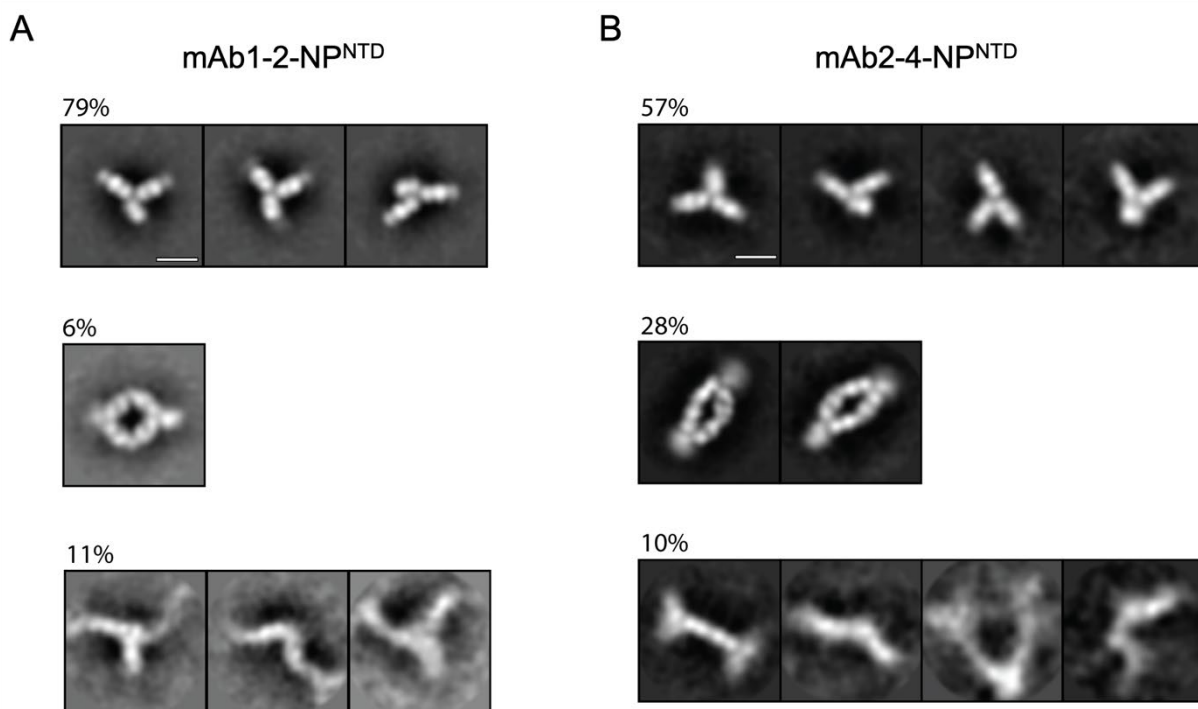
Figure 7-30. NS-TEM 2D class averages with higher concentration and faster dilution mAb-NP^{NTD} samples show distinct pairing differences. Class averaged 2D projections for mAb1-2-NP^{NTD} A), and mAb2-4-NP^{NTD} B).
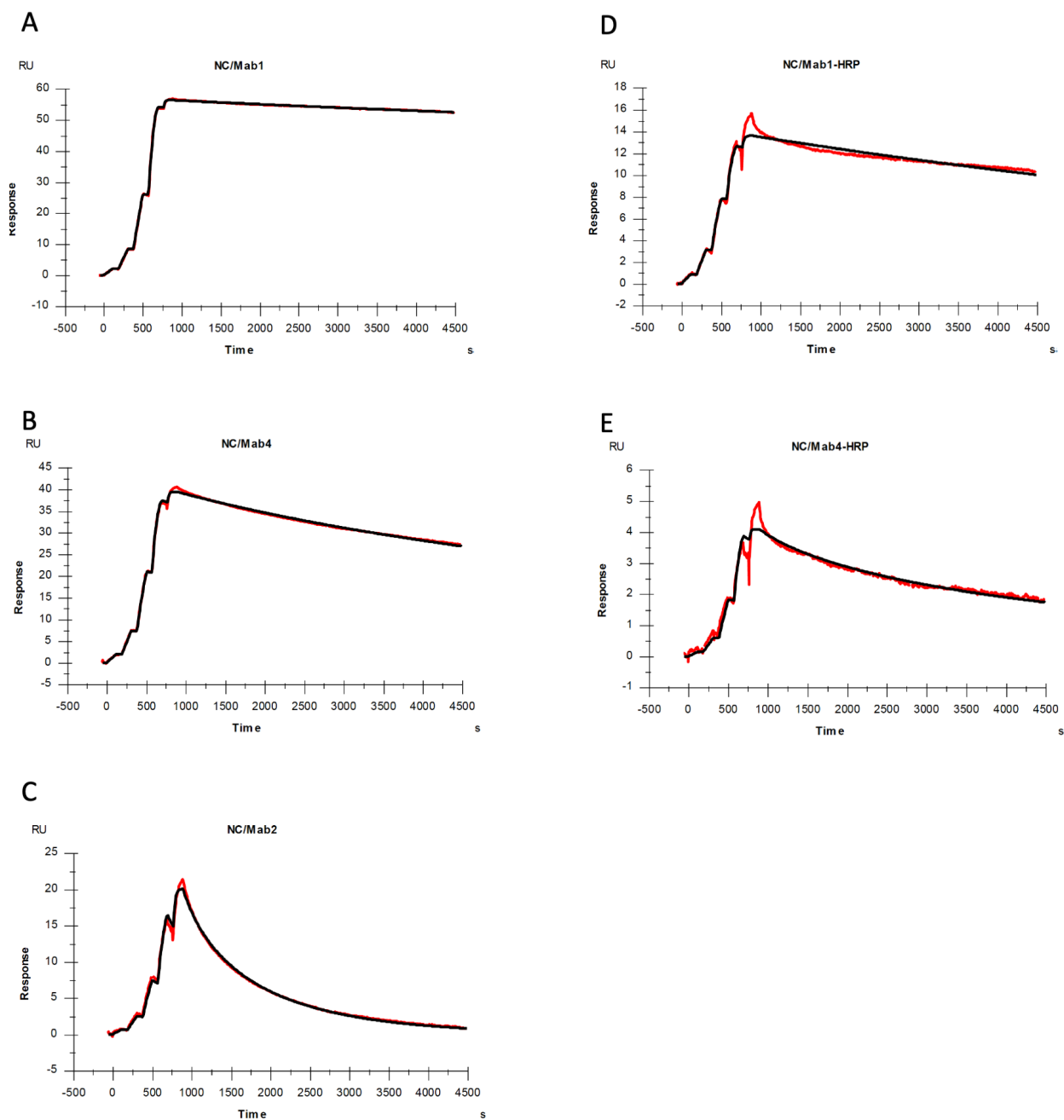
Figure 7-31. Binding kinetics of NP^{NTD} with the mAb 1, 1-HRP, 2, 4, and 4-HRP by surface plasmon resonance (SPR). NP^{NTD} at five concentrations (0.074nM, .22nM, .67nM, 2nM, 6nM) were tested using single-cycle kinetics with immobilized, mAb1 (A), mAb4 (B), mAb2 (C), mAb1-HRP (D), and mAb4-HRP (E).
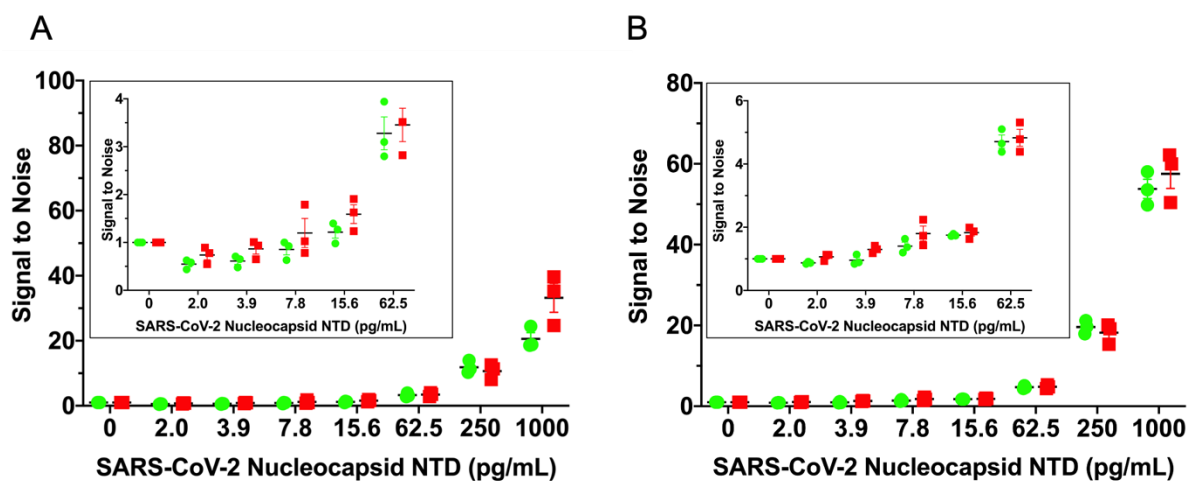
Figure 7-32. Standard ELISA control protocol shows no boost. (A) A standard ELISA where the plate was washed prior to addition of the detection HRP-conjugated mAbs (1-HRP in green, 4-HRP in red), for a twenty minute incubation. (B) A repeat of the experiment shown in A, except the mAb-HRPs had a longer incubation time of 1.5 hrs. In both experiments, no boost is observed, regardless of the mAb-HRP incubation time. Each NP^{NTD} concentration was run in triplicate, and the standard error of the mean for each is included.
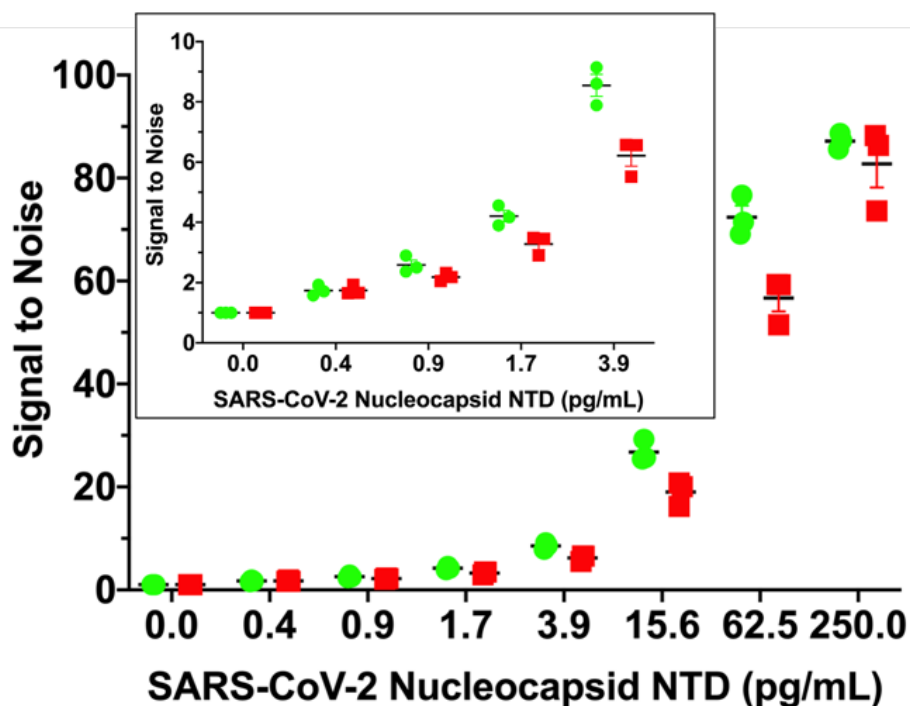
Figure 7-33. The addition of virion lysing triton X-100 increases the LOD of the modified ELISA, diminishing the signal's boost. (A) A modified ELISA where the detection HRP-conjugated mAbs (1-HRP in green, 4-HRP in red) are added directly on top of the samples during the NP^{NTD} capture incubation period, and free (non-plate-bound) mAb2 is "spiked" into the detection HRP-conjugated mAb solutions before their addition on top of the samples. The NP^{NTD} samples were diluted in PBS pH 7.4 plus 0.5% triton X-100. Each NP^{NTD} concentration was run in triplicate, and the standard error of the mean for each is included.

Table 7-6. SPR experimental parameters

Antibody	Association rate constant, k_a ($M^{-1}s^{-1}$)	Dissociation rate constant, k_d (s^{-1})	Dissociation equilibrium constant, K_D (pM)	Experimental Rmax (RU)	Theoretical Rmax (RU)	% Activity of mAb
mAb 1	1.7×10^7	2.2×10^{-5}	1.3	56.4	65.0	87
mAb 1-HRP	7.6×10^6	8.5×10^{-5}	11.0	13.6	45.6	30
mAb 4	1.1×10^7	1.2×10^{-4}	11.0	39.4	43.4	91
mAb 4-HRP	1.7×10^7	4.7×10^{-4}	28.0	4.1	40.6	10
mAb 2	9.4×10^6	1.8×10^{-3}	190.0	20.7	39	53

7.4.5 *Materials and Methods*

7.4.5.1 Monoclonal Antibody Sources

Seven antibodies were purchased from SinoBiological, and two antibodies from CreativeBiolabs. The catalog numbers for the SinoBiological antibodies are as follows: #1: 40143-MM05, #2: 40143-MM08, #3: 40143-R001, #4: 40143-R004, #5: 40143-R019, #6: 40143-R040, #15: 40588-R0004. The catalog numbers for the CreativeBiolabs antibodies are as follows: #18: MRO-0015YJ, #19: MRO-0016YJ. Antibodies #1 and #4 were chemically conjugated to HRP by SinoBiological CRO Services. The molar HRP:Ab ratio was 2.81 for #1-HRP, and 3.5 for #4-HRP.

7.4.5.2 Expression and purification of NP^{NTD}

Gene fragment coding of nucleocapsid protein from SARS-CoV-2 was codon-optimized for efficient expression in *E. coli*. The coding sequence of NP^{NTD} comprising residues Asn47 to Ala173 (UniProtKB - P0DTC9) was synthesized and cloned into pMCSG53 vector [446] by Twist Biosciences, USA. Cloning into pMCSG53 vector introduced to NP^{NTD}, a His$_6$-Tag at the N-terminus followed by a cleavage site for tobacco etch virus (TEV) protease. For NP^{NTD} expression, the plasmid was transformed into *E. coli* BL21(DE3)-Gold cells (Stratagene) using heat-shock. After transformation, bacterial cells were precultured overnight at 37°C in 100 ml of LB Lennox medium supplemented with 40 mM K$_2$HPO$_4$ and 160 mg/L of ampicillin. Subsequently, 40 ml of overnight cultures were used to inoculate 4 liters of LB with 40 mM K$_2$HPO$_4$ and 160 mg/L ampicillin. Next, cells were incubated at 37°C with 180 RPM shaking for approximately 3 hours until reaching optical density at 600 nm equal to 1. Subsequently, the bacteria culture was cooled down for 1 hour in an incubator set to 4°C with 180 RPM shaking.

Expression of NP^{NTD} was induced with 0.2 mM isopropyl β-d-1-thiogalactopyranoside, supplemented with 0.1% glucose, and incubated overnight at 16°C. Bacteria cells were harvested by centrifugation at 4°C, 5000 RCF for 10 minutes. Cell pellets were resuspended in lysis buffer 50 mM HEPES pH 8.0, 500 mM NaCl, 5% v/v glycerol, 20 mM imidazole, and 10 mM β-mercaptoethanol (1 g of cells: 5 ml of lysis buffer) for purification or frozen and stored at -80°C until purification.

After overexpression of NP^{NTD}, bacteria cells were lysed by sonication on ice using 120W output power for 5 minutes (4 sec pulses of sonication followed by 20 sec brakes). After sonication, samples were centrifuged to remove cellular debris (30k RCF, 4°C, 1 hour). We used a vacuum-assisted purification system to perform NP^{NTD} purification with immobilized metal affinity chromatography (IMAC). Using 5 ml of Ni^{2+} Sepharose (GE Healthcare) loaded on a Flex-Column (420400-2510) attached to a Vac-Man vacuum system (Promega), beads were equilibrated in a lysis buffer. The cell lysate was loaded on the column, and Ni^{2+} Sepharose was washed using 20 column volumes of lysis buffer. For elution, lysis buffer was supplemented with imidazole up to 500 mM (pH 8). After elution for removing 6His-Tag, we used TEV protease added in a molar ratio 1 TEV to 40 NP^{NTD}. TEV cleavage leaves 3 residues SerAsnAla at the N-terminus of NP^{NTD}. Next, NP^{NTD} was concentrated using 10 kDa cut-off centrifugal protein concentrators (Merck-Millipore). Subsequently, we performed SEC of NP^{NTD} using a Superdex S200 16/600 column attached to an Äkta Express (GE Healthcare) purification system. SEC was done at 4°C in a buffer containing 20 mM HEPES, 500 mM NaCl, 5% v/v glycerol, 10 mM β-mercaptoethanol, pH 8.0. Purified fractions of NP^{NTD} from the middle of the gel filtration elution peak were concentrated to 10.7 mg/ml. Protein was flash cooled using 40 μL aliquots dropped directly into liquid nitrogen. Samples were stored at -80°C or on dry ice during shipment. Upon thawing, samples were stored at 4°C and diluted in PBS pH 7.4 for functional assays.

7.4.5.3   Size exclusion chromatography coupled to small angle X-ray scattering with multi-angle light scattering

For SEC-MALS-SAXS experiments, 60 μL of samples containing mAb ~1-3 mg/mL and NP^{NTD} in 1:5 molar ratio were prepared in PBS pH 7.4 buffer. mAb 1, 2, and 4 were also measured in the absence of NP^{NTD} using the same buffer conditions. The mAb pairs 1-2, 1-4, and 2-4 in the presence of NP^{NTD} were prepared in the molar ratio of 1:1:10 in the same buffer conditions. All samples were incubated for a minimum of 30 minutes before the injection on SEC.

SEC-MALS-SAXS was collected at the SIBLYS beamline (BL 12.3.1) at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL) in Berkeley, California. [333] X-ray wavelength was set at λ=1.216 Å, and the sample to detector distance was 2070 mm, resulting in scattering vectors, q, ranging from 0.01 Å^{-1} to 0.4 Å^{-1}. The scattering vector is defined as q = 4πsinθ/λ, where 2θ is the scattering angle. All experiments were performed at 20°C, and data were processed as previously described. [9] Briefly, a SAXS flow cell (see Section 2.3.2) was coupled with an inline Agilent 1290 Infinity HPLC system using a Shodex KW 803 column. The column was equilibrated with running buffer (PBS pH 7.4) with a 0.65 mL/min flow rate. 55 μL of each sample was run through the SEC, and two second x-ray exposures were collected continuously during a 20 min elution. The SAXS frames recorded before the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by

the radius of gyration (Rg) derived by the Guinier approximation $I(q) = I(0) \exp(-q^2 * Rg * 2/3)$ with the limits $q * Rg < 1.5$. The elution peak was mapped by comparing integral ratios to background and Rg relative to the recorded frame using the program SCÅTTER. Uniform Rg values across an elution peak represent a homogenous state of mAb or its complex. Final merged SAXS profiles (Figure 7-27), derived by integrating multiple frames across the elution peak, were used for further analysis, including a Guinier plot, which determined the aggregation free state (Figure 7-27). The program SCATTER was used to compute the pair distribution function (P(r)) (Figure 7-22B and Figure 7-24C). The distance r where P(r) approaches zero intensity identifies the macromolecule's maximal dimension (Dmax, Figure 7-22B, Table 7-5). P(r) functions for single mAb (Figure 7-23A), single mAb+NP^{NTD} (Figure 7-22A) were normalized at the maxima except the P(r) of NP^{NTD} alone (Figure 7-22B); mAb1-2-NP^{NTD} and mAb2-4-NP^{NTD} complexes (Figure 7-24B) where the area under P(r) function correlates to MW estimated by SAXS. The SAXS flow-cell was also connected inline to a 1290 series UV-vis diode array detector measuring at 280 and 260 nm, 18-angle DAWN HELEOS II multi-angle light scattering (MALS) and quasi-elastic light scattering (Wyatt Technology), and Optilab rEX refractometer (Wyatt Technology). System normalization and calibration were performed with bovine serum albumin (BSA) using a 45 μL sample at 10 mg/mL in the same SEC running buffer and a dn/dc value of 0.175. The light scattering experiments were used to perform analytical scale chromatographic separations for MW and hydrodynamic radius (Rh) determination. UV, MALS, and differential refractive index data were analyzed using Wyatt Astra 7 software to monitor sample homogeneity across the elution peak complementary to the above-mentioned SEC-SAXS signal validation.

7.4.5.4   Solution state modeling

BILBOMD [431] rigid body modeling along with a FoXS and MultiFOXS [284,291] approach was used to define, select and weight the two-state atomistic model that best agreed with individual SAXS profiles of free mAb1, 2 and 4. The crystal structure with PDB ID: 1HZH, [447] including the glycans moiety, was used as an initial model. In the case of glycoproteins, the glycans' contribution to the scattering is known to be larger than protein alone. [448] Minimal molecular dynamics simulation applied on the mAb hinge regions explores the Fab domain's conformational space relative to the Fc-glycan region. The disulfide bonds in the hinge region were kept intact. In the conformational sampling, individual Fabs would move independently of one another. A single best-fit two-state and three-state model was selected for each mAb using MultiFOXS. [284,291] We estimated the number of mAb states in solution by examining the R_g distribution [282] for the top 300 multistate-models (Figure 7-23D) that all gave the same goodness of fit (χ^2 ~0.9). The number of main peaks in the distribution indicates the number of states. The area and the position of the peaks validate the level of mAb flexibility.[277,282] The SAXS envelopes were reconstructed from the experimental data of mAb2-4-NP^{NTD} and mAb1-2-NP^{NTD} complex using the program DAMMIF. [449] Ten bead models obtained for each SAXS experiment were averaged by DAMAVER [450] to construct the average model representing each reconstruction's general structural features. Bead models were converted to volumetric SITUS format with the pdb2vol kernel convolution utility. [451] SAXS data and SAXS-derived models

have been deposited in SIMPLE SAXS database (https://simple-saxs.herokuapp.com/), and experimental SAXS parameters are reported in Table 7-5.

7.4.5.5 Negative staining and transmission electron microscopy (TEM)

Carbon film-coated 400 mesh Cu grids [452] were rendered hydrophilic by processing in a Tergeo-EM plasma cleaner (PIE-Scientific) for 30s at 15W. Each complex composition was dialyzed and diluted to a final "low" concentration of 45 nM (2 hr) or "high" concentration of 450 nM (10s) into Tris pH 7.4, 150 mM NaCl buffer from the initial 45 µM condition used for SAXS measurements to prepare NS-TEM specimen grids. Four µL of each diluted specimen were incubated on the pre-treated grids for 1 min before staining in 4 successive 50 µL drops of 1% uranyl formate on Parafilm, then blotted and air dried. [453] The grids were imaged in a Tecnai F20 electron microscope (ThermoFisher Scientific) operated at 120kV and equipped with a DirectView direct electron detector (Direct Electron) with a pixel size of 1.3 Å/pixel, at a cumulated dose of 40 electrons/pixel and a defocus range of -0.6 to -1.3 µm.

7.4.5.6 NS-TEM data processing

For each specimen, single particle data were picked manually from the micrographs and processed in RELION-3.1. [454,455] The final data sets added up to 539, 6548, 693 and 3427 particles from 39, 82, 38 and 59 micrographs for mAb1-2-NP^{NTD} "low" and "high" concentration, and mAb2-4-NP^{NTD} "low" and "high" concentrations, respectively. Contrast transfer function (CTF) was estimated, particle data was extracted as 128x128 or 152x152 pixel boxes, 2.7 Å/pix, CTF-corrected and normalized before undergoing 2D alignment and classification. Class-average features were interpreted based on the number and arrangement of identifiable antibodies in their densities. Corresponding classes were grouped accordingly, and their populations were pooled.

7.4.5.7 SPR

Affinity and kinetic data were acquired using a Biacore T200. All antibodies were coupled to a CM5 Biacore sensor chip using amine coupling. MAbs 1, 2, 4, 4-HRP were coupled using 10 mM acetate pH 5.5 at 10 ug/mL, while mAb 1-HRP required 10 mM acetate pH 5.0 at 10 ug/mL. All experiments were run in 10 mM HEPES pH7.4, 150 mM NaCl, 35 mM EDTA, 0.01% surfactant P20. NP^{NTD} analyte injections for the single-cycle kinetic titrations were as follows: 0.074 nM, 0.22 nM, 0.67 nM, 2 nM, 6 nM. The dissociation time was 3600s.

7.4.5.8 ELISA development

ELISAs were developed using the following materials: Corning 96-Well High-Binding Flat-Bottom Microplates from StemCell (Cat. # 38019), and R&D Systems, Stop Solution 2N Sulfuric Acid (Cat. # DY994), Substrate Reagent Pack (Cat. # DY999), BSA-ELISA grade (5217/100G). All reagents were allowed to warm to room temperature before use. ELISA signals were recorded using a POLARstar Omega plate reader. Samples were diluted in PBS unless

otherwise stated. The wash buffer consisted of PBS pH 7.4 and 0.05% Tween 20. The blocking buffer consisted of PBS pH 7.4, 2% BSA, and 0.1% Tween 20. Unless specified, the following modified ELISA protocol was used: 100 µL per well of 4 mg/mL mAb2 capture antibody in PBS was added to the ELISA microtiter plate. The plate was sealed and incubated overnight at room temperature. The next day, the solution was discarded and washed with wash buffer three times. The plate was blocked with 300 µL per well of blocking buffer for one hour at room temperature. The plate was washed three times. NP^{NTD} was diluted into PBS as a serial dilution concentration series, and 100 µL per well was added to the plate. For detection and "antibody networking" assessment, 100 µL of PBS containing 0.4 mg/mL antibody-HRP with or without 0.4 mg/mL mAb2 capture antibody was added directly to the samples in the plate, for a final concentration of 0.2 mg/mL. The plate frame was gently tapped for one minute to mix, sealed, and incubated protected from light for 1.5 hours at room temperature. The plate was washed three times. Before use, the Substrate Reagent was prepared by combining equal parts of component A & B. 100 µL of working Substrate Reagent was added to each well and incubated for 20 minutes at room temperature protected from light. 50 µL of Stop Solution was added to each well, and the plate was gently tapped to ensure thorough mixing. The optical density (OD) of each well was determined within 30 minutes of stopping the reaction. The OD 450nm and 540nm were recorded. The data was background corrected in excel by subtracting the OD 540nm from the 450nm signal. For normalized data, all signals were individually divided by the background signal. All samples were run in triplicate. The mean and standard error of the mean was calculated and plotted in GraphPad Prism.

7.5 Case 4: Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication***

7.5.1 Introduction

The SARS-CoV-2 virus has plagued every populated continent and has been implicated in 3,000,000 deaths at the time of writing. Recently deployed vaccines have provided much hope. However, distribution, voluntary immunization and mutations in the virus remain major concerns. The development and application of treatments that attack fundamental viral machinery will remain critical for the foreseeable future.

In the US, thus far, a single drug Remdesivir has been conditionally approved for specifically targeting the virus. It targets the RNA-dependent, RNA polymerase (RdRp) [456]. SARS-CoV-2 RdRp, also referred to as nonstructural protein 12 (Nsp12), is the catalytic center of the macromolecular complex frequently referred to as the replication-transcription complex (RTC). The RTC is essential for virus replication as it makes copies of genomic and subgenomic RNAs and polymerizes anti-sense RNA. After maturation, these RNAs serve as mRNAs to produce virus Nsps, structural and accessory proteins. Copies of genomic RNA are eventually packaged into mature virion that bud out of host cells. The Remdesivier triphosphate is incorporated into newly synthesized RNA by RdRp and either stalls or interrupts viral RNA polymerization.

While the corona-virus genome is larger than most viruses (~30kbs), it remains 100,000 times smaller than that of humans. The highly evolved and efficient SARS-CoV2 RTC is a major part of how the virus overcomes the limitations of its genome size. The RTC specifically identifies and polymerizes viral over host RNA without activating host defenses. The RTC contextually balances synthesis of subgenomic RNAs for use in translation of viral proteins and polymerization of genomic RNA for new virus maturation. The subgenomic RNA produced must be 3' polyadenylated and 5' capped conforming to host ribosome mRNA conventions. The RTC can discontinuously read long RNAs and produce shorter subgenomic RNAs that serve as mRNA that code for structural and accessory proteins within the long RNA (discontinuous transcription) [457]. The mechanisms must be robust to varied attacks by the host cell defenses. All this while retaining primary polymerase activities of recognizing initiation and termination sequences, discriminating DNA vs RNA basis, and base pairing. Reverse engineering this complex macromolecule is likely to be a challenge. However, as a central mechanism with so many functions, several avenues for sabotage by small molecules or proteins may be possible beyond the one exploited by Remdesivir.

In SARS-CoV-2, besides RdRp (Nsp12), the RTC involves non-structural proteins Nsp7, Nsp8 and others [458-460]. For an excellent review that largely summarized our understanding prior to the pandemic see Snijder et al [461]. Nsp12 is the largest component of the complex while Nsp7 and Nsp8 are considered co-factors.

Several additional functions have been reported for Nsp7 and 8. Both Nsp7 and 8 are transcribed at higher rates relative to Nsp12 as part of ORF1a and earlier in the infection so they may act to prepare the cell for viral replication. In a thorough study, Nsp8 has been shown to polyadenylate RNA [462] which could play a role in 3' mRNA preparation. Nsp8 has also been implicated in blocking ribosomal membrane protein recognition signals [463]. Crystal structures and other supporting methods have visualized Nsp7/8 in a ring like hexadecamer [464] that provocatively suggests a processive sliding complex on double stranded RNA with possibly independent role.
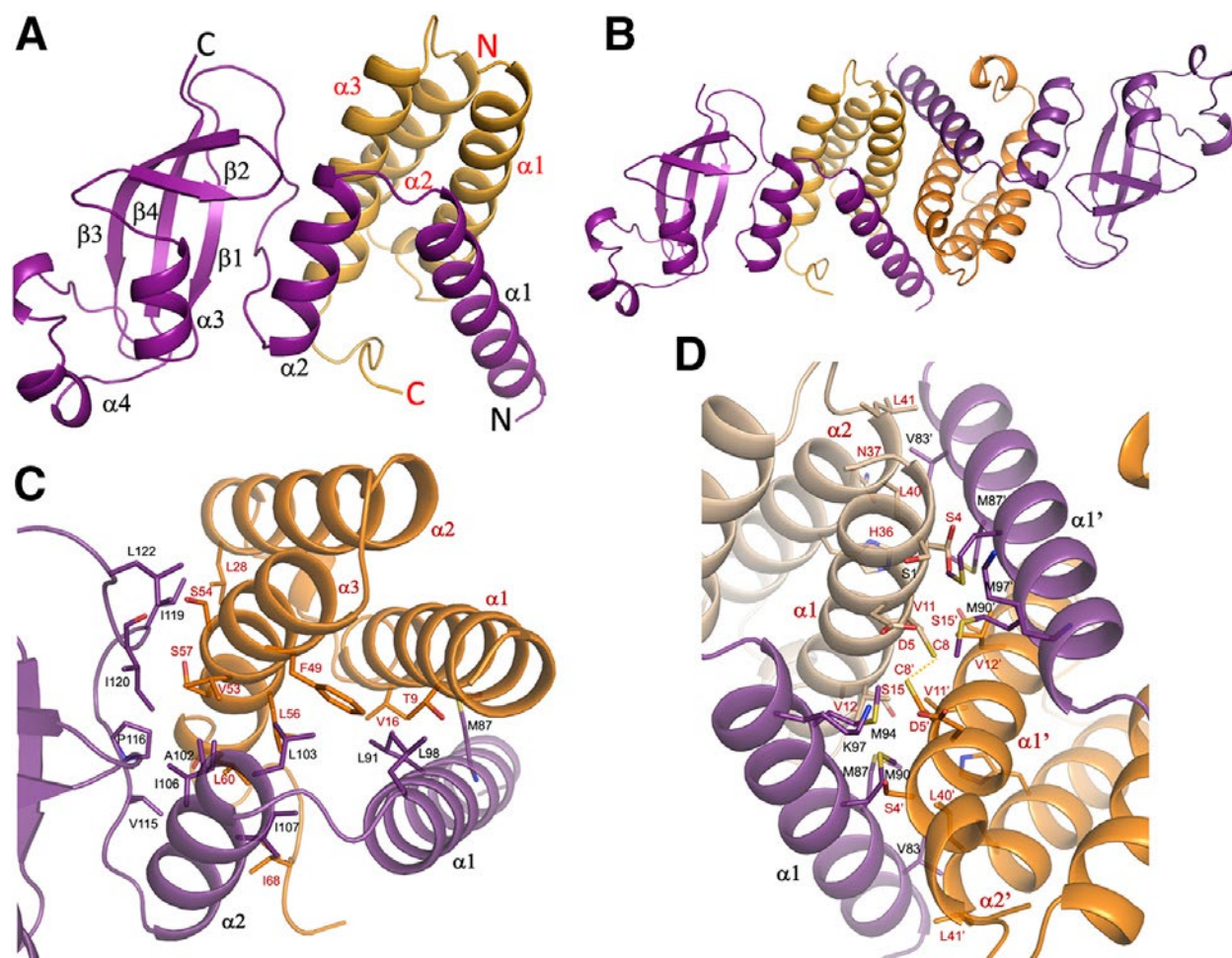
Figure 7-34. Structure of Nsp7/8 complex. The crystallized structure of the heterodimer Nsp7/8 (A) shows Nsp7's (orange) C-terminal helices intercalated between Nsp8's (purple) long a1 N-terminal helix (truncated in our structure) and a2. The heterotetramer structure (B) is also shown and present in all three of our crystal forms. Details of the heterodimer interface are shown in (C). Details of the heterotetramer interface are shown in (D), with interacting residues shown by sticks labeled red for Nsp7 and labeled black for Nsp8. An interchain, symmetrically formed disulfide bond in Nsp7 formed by C8 is shown in dashes.

Our understanding of the RTC has been remarkably advanced through recent cryo-EM studies resolving structures of Nsp7/8/12 with and without RNA [456,459,465,466]. Most recently, structures of Nsp7/8/12/13 [467] [468] and Nsp7/8/9/12/13 [469] have been reported. However, many questions remain about connecting the cryo-EM structures with the assembly, role of co-factors and mechanism. The integration of many structural, biochemical and genomics studies will be required to provide a mechanistic and actionable model of each protein contributing to the RTC; to assess their assembly, varied functions and potential vulnerabilities.

Toward the goal of understanding the macromolecular machinery of the RTC we have undertaken a systematic study of individual components and their complexes. We have solved a 1.5 Å X-ray macromolecular crystal structure of the complex between Nsp7 and Nsp8 (Figure 7-34). This structure along with those solved by others [470], particularly the recent cryo-EM structures of the Nsp7/8/12 RNA complex [471], provided key information for the interpretation of our solution scattering measurements. We have used solution scattering along with other biochemical techniques to study each protein in isolation, pairs (Nsp7/8, Nsp8/12, Nsp7 RNA, and Nsp8 RNA), ternary complexes, and all four components together (Nsp7/8/12 RNA) (Fig 2, 4, 5, and 6).

As many proteins in the viral genome are expected to have multiple functions that require conformational or assembly changes, heterogeneities are expected in solution. All complexes were examined during elution from a size exclusion purification column (SEC) by multi-angle light scattering (MALS) and small angle X-ray scattering (SAXS); SEC-MALS-SAXS. MALS provides mass and SAXS informs on structure at every point in elution. Together they detail the oligomeric and conformational heterogeneity characteristic of each mixture, complementing the cryo-EM and crystallographic results. To further interpret complexes, we also utilized small angle neutron scattering (SANS) and its capability to contrast match out components within complexes. Using these and other biochemical techniques we provide insights into function, the assembly process, dynamics and identify stable complexes that are tractable for further mechanistic interrogation.

7.5.2   Results

7.5.2.1   Atomic Resolution Crystal Structure of Nsp7/8

To provide insights into the structures, key residues of association and multimeric state between the SARS-CoV-2 RTC's co-factors, we solved high resolution crystal structures of Nsp7 bound to Nsp8. Structures were determined in three different crystal forms, Nsp7/8A, Nsp7/8B, and Nsp7/8C and deposited in the protein data bank (PDBIDs: 6WIQ, 6WQD and 6XIP, Table 7-7) early in the pandemic. Our deposited structures are similar to that reported later by Konkolova et.al[470]. Crystals of Nsp7/8A were shown to have a truncated N-terminus (at residue Glu78) for Nsp8. The other structures were obtained from crystallization in the presence of protease Glu-C (V8). Below, the description of the Nsp7/8 structure is based on the 1.5 Å structure Nsp7/8C, unless otherwise mentioned.
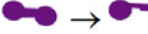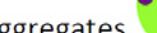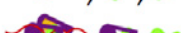
Macromolecular Mix	Number of SEC-MALS-SAXS peaks	Apparent mass of peak(s), (kDa)	Model that fits SAXS in peak
Nsp7	2 (Fig S2)	>30, **17**	
Nsp8	1 (Fig 4A)	40 → 28	
Nsp12	1 (Fig 4A)	105	
ssRNA	1 (Fig 4A)	~20	
dsRNA	1 (Fig 4A)	~15	
Nsp7/8	2 (Fig 4A)	**50 →36**, 17	
Nsp8/12	2 (Fig 5A)	>500, **125**	Aggregates,
Nsp7/8/12	3 (Fig 5A)	**125**, 100, 17	
Nsp7 dsRNA	2 (Fig S2)	~15, 17	
Nsp8 ssRNA	1 (Fig 4A)	45→35	
Nsp8 dsRNA	2 (Fig 4A)	45,**45→20**	
Nsp7/8/12 ssRNA	3 (Fig 5A)	**190**, ~120, ~100	
Nsp7/8/12 dsRNA	3 (Fig 5A)	**195**, ~120, ~100	

Figure 7-35. Nsp7, Nsp8, Nsp12, and RNA complexes were found in elution profiles from SEC-MALS-SAXS. Each molecule was measured independently and is represented pictorially (top left). The number of peaks and the figure associated with the elution profile is indicated in the second column. The apparent mass through each peak is indicated in the third column. When an elution has more than one peak, the largest peak is indicated by bold mass. In some peaks, the mass changes across the peak. Mass values prefaced with ~ indicate weak signals by MALS because of low abundance. For comparison, the calculated masses for monomeric Nsp7, Nsp8, and Nsp12 are 9, 24, and 100 kDa respectively. The rightmost column depicts mixtures of models that fit the SAXS data through analysis described in the remainder of the text.

The structures of Nsp7 and Nsp8 have strong similarities to the same proteins and variants from other coronaviruses. Nsp7 (Figure 7-34A) has three consecutive α helices (α1, α2 and α3) forming a three helical coiled coil bundle and the C-terminal loop (residues 62-70) that in some structures is followed by a short not well-defined helix (residues 68-72). The conformation of Nsp8 has previously been described to resemble a golf-club[464] (Fig 1A). It has an N-terminal α-helix (α1) which starts in our structure at Asp78 due to the truncation by Glu-C, which is highly positively charged for binding RNA. This helix is followed by the second α-helix (α2) that connects through a long loop to a half β-barrel-like domain formed by five antiparallel β strands (β1-5) with a small α-helix (α3) inserted between the first and second strands and another insertion of a long loop that contains two half-turn helices (labeled α4). The C-terminus (193-198) of the Nsp8 is well defined in Nsp7/8B but not in other forms.

In all of our crystal forms, the heterodimer formed between Nsp7 and 8 places Nsp7 near the half β-barrel-like domain of Nsp8 with the N-terminal helix of Nsp8 extended and pointing away (approximate dimensions of 40 x 40 x 46 Å) (Figure 7-34A). The similarity between all heterodimers in all our crystal forms is remarkable with an RMSD of 0.82 Å for 184 aligned Cα atoms despite different unit cell dimensions, different space groups and different asymmetric units. These heterodimers are also similar to those from other coronaviruses, for example, RMSD between the heterodimer of Nsp7/8C and that of the Nsp7/8 from SARS CoV (PDBID 5F22) is 0.91 Å for 184 Cα atoms. These similarities suggest that, once formed, the heterodimer is rigid.

The interface between Nsp7 and Nsp8 is made through hydrophobic contacts between six bundled helices. The helices involved are the two N-terminal helices (α1 and α2) from Nsp8 and four helices from Nsp7. At the C-terminus, the fourth helix in Nsp7 is not well defined and the loop of residues 66-72 wedges in between the two (α1 and α2) Nsp8 helices to extend the interacting surface. The dimer interface area is 2834 Å^2 with 72 % hydrophobic contacts. The key interfacial residues are shown in Figure 7-34C.

Our heterodimer structure is in good agreement with the recent cryo-EM structures of Nsp7 and Nsp8 bound to Nsp12 with and without RNA. Using the most complete structure (PDBID 6YYT) from Hillen et al, the superposition has an RMSD of 1.0 Å. There are two Nsp8s in this structure but only one is in contact with Nsp7. The interactions between Nsp12 and the Nsp7/8 are mostly through the Nsp7, and the region known as the thumb domain (residues 812-932) located at the C-terminus of Nsp12. The Nsp8 monomer interacts with the fingers domain (residues 250-398) of Nsp12 primarily through α2, the following loop and β1 (residues 99-126) and has a significantly different conformation compared to the Nsp7/8 heterodimer. The C-terminal two thirds of Nsp8 (residues 127-192) is quite similar (overall difference of 1.2 Å over 67 residues). Finally, the location and orientation of the N-terminal helices are different from those in the Nsp7/8 heterodimer. In the cryo EM structures the N-terminal helix of Nsp8 makes significant contacts with Nsp12 including the Glu81 region which perhaps protects Nsp8 from the proteolysis we observed during our crystallographic efforts.

In contrast to previously reported Nsp7/8 structures from other coronaviruses, in all three of our structures, either as in the asymmetric content or symmetry-related, a hetero-tetramer or a dimer of heterodimers is present as an elongated, thick rod shape of 40 x 92 Å (Figure 7-34B). The differing crystal forms show only slight variation in the packing of Nsp7/8 heterodimers to form hetero-tetramers. Comparing the hetero-tetramers in the asymmetric units, the RMSD is 1.28 Å for 368 aligned Cα atoms. The hetero-tetramers are formed between the two heterodimers that are related by the non-crystallographic two-fold rotational symmetry. In Nsp7/8A, the tetramers are formed by the crystallographic two-fold rotational symmetry.

The hetero-tetramer interface involves a total of six helices: α1 and α2 of Nsp7 and their symmetry-mates (two-fold rotation) α1' and α2' and α1 of Nsp8 and the symmetry-mate α1'. The symmetry related secondary structures are indicated by apostrophes. The tetramer interface is not as extensive as that of the dimer having 1813 Å^2 with 73% being hydrophobic. The residues involved are mostly making hydrophobic contacts with symmetrically related residues. Importantly, a symmetrically related inter-chain disulfide bond is also found between Cys8 of Nsp7 and its symmetry mate (Figure 7-34D).
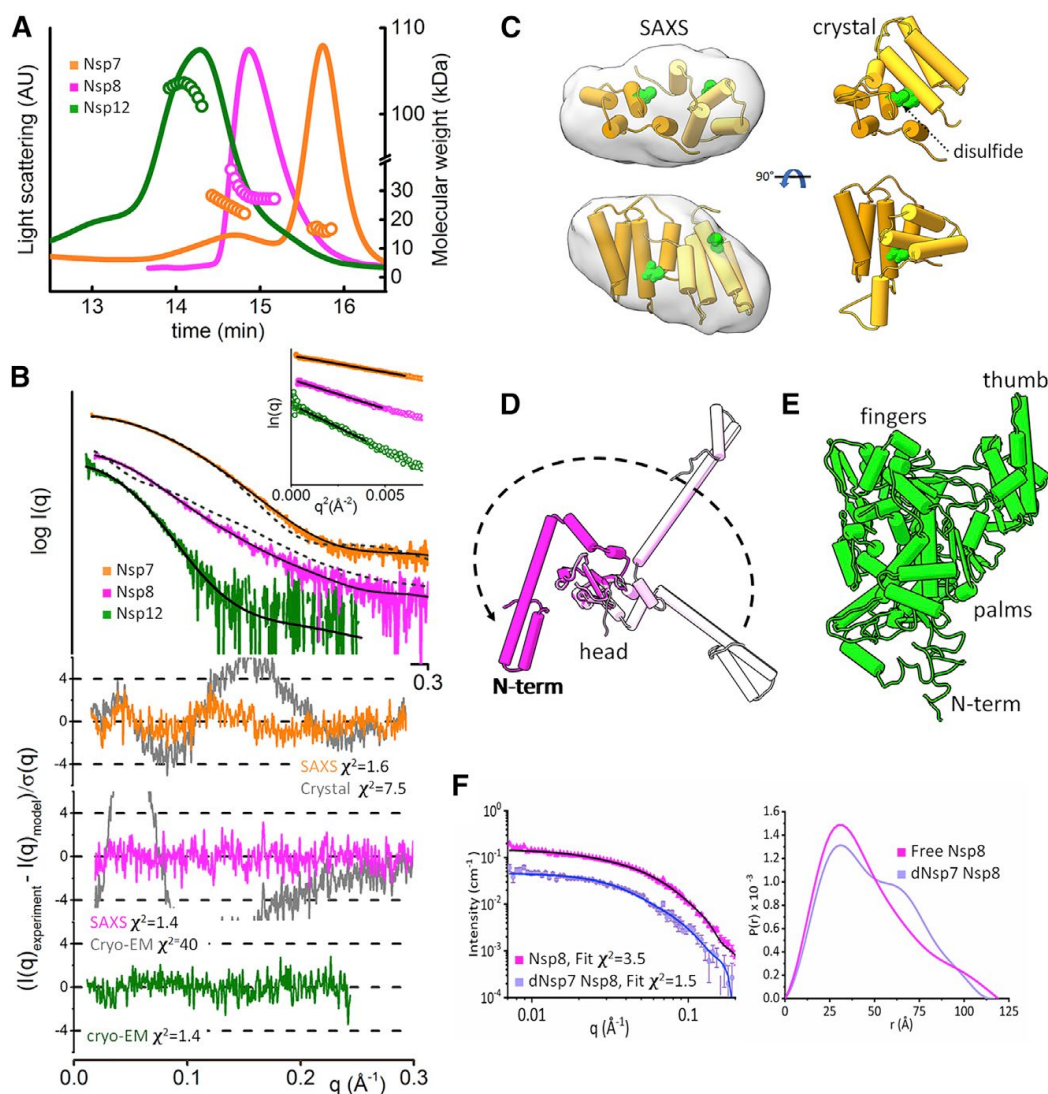
Figure 7-36. Solution states of independent Nsp7, Nsp8, and Nsp12 (A) SEC-MALS-SAXS elution profiles for Nsp7 (orange), Nsp8 (purple), and Nsp12 (green) by light scattering intensity (solid lines, left axis), with mass indicated by circles (right axis). (B) Experimental (colored lines) SAXS profiles for each protein. Guinier plots for experimental SAXS curves are shown in the inset. Calculated best-fit models (solid black lines) and alternate models from available structures (dashed lines) are shown along with residuals (lower plot, gray for alternate models) and goodness-of-fit parameter c2. (C) Best-fit model for Nsp7 is an alternate dimer than that found in our crystal structure, with the disulfide forming Cys8 shown in green. The average SAXS envelope is superimposed on the SAXS model. (D) The Nsp8 monomer is found in a thus far unobserved conformation (dark magenta) relative to the N-terminal (N-term) domain in the superimposed atomically resolved cryo-EM structures (pink) (PDB: 6YYT). (E) Nsp12 measurements agree with available atomic structures (PDB: 6YYT). (F) SANS profiles (left) were measured for Nsp8 (magenta) and dNsp7-Nsp8 complex (light blue circles) in 90% D2O, masking dNsp7. Fits to models described in the text are shown in black and blue, respectively. The P(r) calculated from SANS for both (right plot) shows Nsp7 alters Nsp8 structure.

234

Among the two crystal structures of Nsp7/8 complexes from SARS-CoV reported, the tetrameric form (PDBID 5F22) is similar to our tetrameric structures. The second form (PDBID 2AHM) of the hexadecameric superstructure of eight each of Nsp7/8 presents a cylindrical ring, ~90 Å long and ~ 30 Å inner diameter lined with positive charge and could accommodate two and a half turns of RNA duplex [464] suggesting a role as a primase for Nsp12. In the hexadecamer, there are two different heterodimers with two different conformations of Nsp8. The tetramer interface is angularly shifted to propagate a ring-like rather than linear topology. In our tetrameric form, each heterodimer is in near identical conformations.

7.5.2.2   Assemblies in Solution by SAXS and SANS

We demonstrated we could successfully produce an active RTC complex from Nsp7/8/12 using a radioactive based label extension assay. The extension assay showed RTC creating duplexed 36 base paired RNA from a 31 base hairpin substrate where the 5' overhang had 5 unpaired bases (Fig S1). Confident in the relevance of our largest complex, we sought to characterize all potential sub-complexes.

To inform on assembly and conformation in solution we collected SEC-MALS-SAXS, SANS and other biochemical data on mixtures of Nsp7, Nsp8, Nsp12 and RNA (Figure 7-35). Prior to experiments all samples were extensively purified and prepared with the sequences as described in Methods Section 7.5.5. We used two primary RNA substrates. Our double stranded RNA (dsRNA) has an 8 base overhang and 28 bases of duplex (Figure 7-42). Uridine composes the first four bases of the 5' overhang. The sequence is analogous to that used in a cryo-EM study [471] differing only in that the strands are continuous rather than containing breaks. The 36-base single-stranded RNA (ssRNA) is derived from a long chain of dsRNA and is referred to as long ssRNA.

Each SEC-MALS-SAXS elution contains considerable information (e.g., Figure 7-36A). Analysis of the MALS on elution traces (left axis, solid and dashed lines) is used to determine the molecular weight (MW) (right axis, circles). Analyzed SAXS profiles are generated by integrating in regions of constant mass and Rg from SEC-MALS-SAXS (e.g., Figure 7-36B) across the main peaks in an elution. Calculations performed on an atomic model generate a SAXS curve (using FOXS [291,336] that can be compared to the experiment using χ^2 or residual as metric of agreement. If agreement is poor (e.g. Figure 7-36B lower plots) then the models are adjusted by molecular dynamics in BILBOMD [283] until an adequate fit and residual are attained. In cases of conformational flexibility, a single model will not be sufficient to fit the data and an ensemble of models must be used (e.g., Figure 7-36C, D, E). All scattering measurements were made in slightly reducing conditions, mimicking intra-cellular space, to prevent disulfide bond formation. Using the isotope-based masking properties of SANS, proteins were also examined for conformational change when interacting with masked components in complexes (e.g., Figure 7-36F). Global parameters extracted from these SAXS and SANS profiles can be found in Table 7-8. Below we organize our results in order of increasing mixture complexity.

7.5.2.3 Nsp7 forms multimers in solution and does not bind RNA.

The SEC-MALS-SAXS of Nsp7 reveals that Nsp7 forms multimers with a different organization than that found in our crystal structure of Nsp7/8 (Figure 7-35 and Figure 7-36). From the dominant late peak, MALS-SAXS measurements are consistent with a well folded dimer with a mass of 17kDa (monomeric mass 9kDa) and a maximal dimension (Dmax) of 60Å (Figure 7-41B). The primary peak in the MALS based chromatogram is preceded with larger multimer of Nsp7 (Figure 7-36A). The presence of larger multimers of Nsp7 are important to note as they complicate the analysis of complexes of Nsp7 with Nsp8, Nsp12, and RNA described further. The Nsp7 multimers can erroneously lead to an interpretation that Nsp7 is part of larger complexes as homo-oligomers of Nsp7 co-elute with larger macromolecular complexes.

Indicative of a different assembly in solution, SAXS calculated from the Nsp7 dimer taken from the Nsp7/8C crystal structure disagreed with the measured SAXS data (χ^2=40). Poor agreement remained (χ^2=7.5, Figure 7-36B) even after optimized remodeling of the C-terminal helix (69-84) that is solvent-exposed in the absence of Nsp8. Furthermore, the shape calculated from our SAXS data by DAMMIF [472] is more elongated than the dimer taken from the Nsp7/8 structure (Figure 7-36C). Assuming the Nsp7 monomer fold, visualized in our crystal structure, orientations of Nsp7 in the dimer that fit the SAXS data were found using the SAXS guided computational docking program FOXDOCK [284]. The best docking results were from head to tail orientations of Nsp7 monomers (χ^2 =1.6) (Figure 7-36C). The head to a tail interface may also allow the formation of larger oligomers observed in MALS chromatograms (Figure 7-41A). These oligomers were further analyzed and were consistent with an elongated rather than globular form of Nsp7 dimers (Figure 7-41B).

Nsp7 did not complex with RNA by SEC-MALS-SAXS analysis. Measurements of mixtures of Nsp7 with RNA resulted in SAXS and masses consistent with the two eluting separately (Figure 7-41A). These results were supported by comparisons to measurements on the RNA run separately (Figure 7-42A, B).

7.5.2.4 Nsp8 forms concentration-dependent multimers and has an alternate conformation when monomeric.

Nsp8 forms concentration-dependent oligomers in solution. SAXS profiles collected in high throughput mode at three concentrations (2.5, 5, 10 mg/ml) varied more significantly than could be explained by a concentration-dependent long-range attraction, indicating complexation between monomeric units (Figure 7-43A). Similarly, SEC-MALS-SAXS from Nsp8 showed an asymmetric elution peak with the front of the peak having a larger molecular mass than the later eluting material that matched the monomeric mass of Nsp8 (Figure 7-36A).
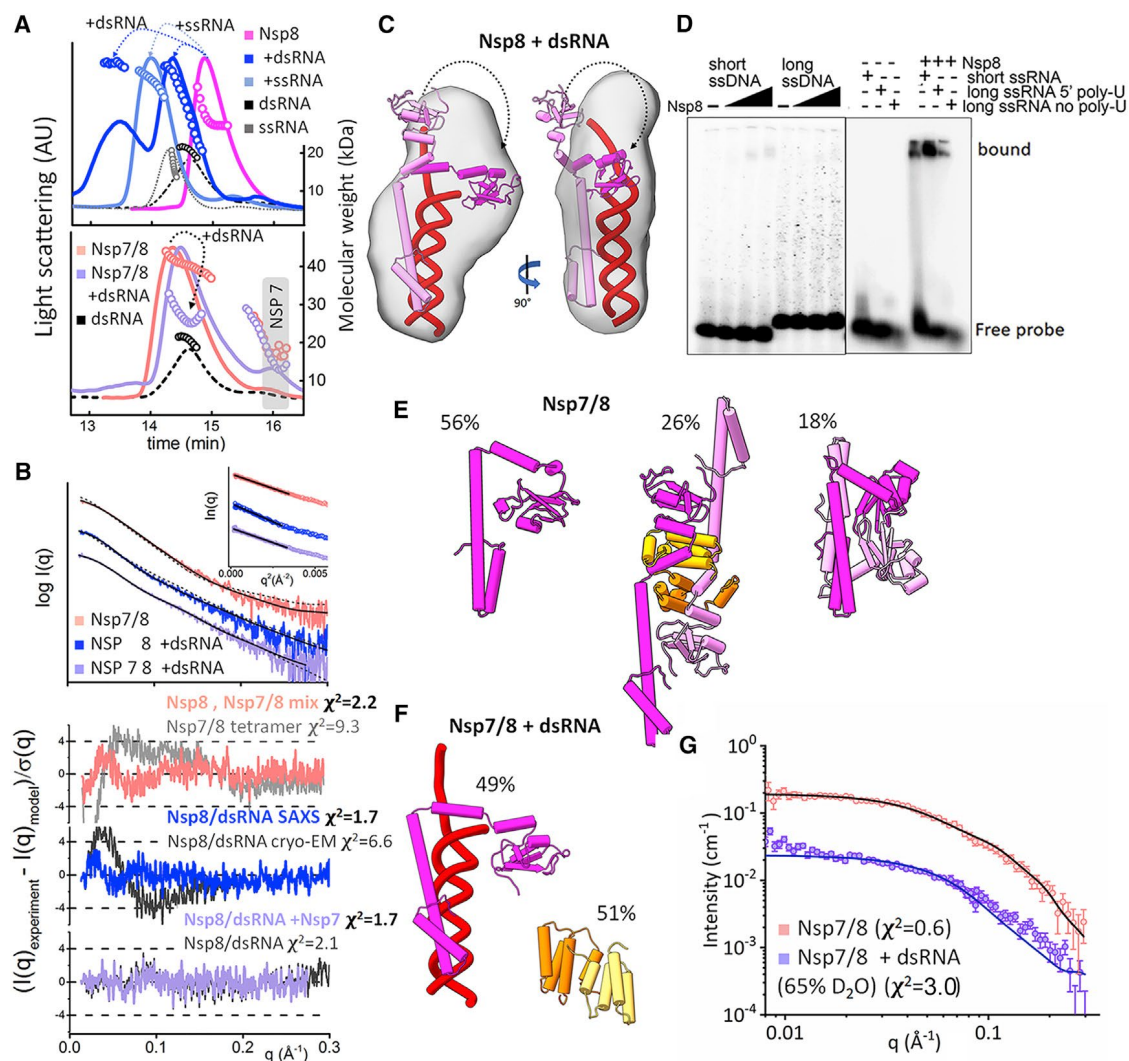
Figure 7-37. Complexation of Nsp8 and Nsp7/8 with nucleic acid. (A) SEC-MALS chromatograms for Nsp8, Nsp8 þ dsRNA, Nsp8 þ ssRNA, dsRNA, and ssRNA (top) and Nsp7/8, Nsp7/8 þ dsRNA, and dsRNA (bottom) are colored as indicated. Solid lines represent the light scattering in detector units (left axis), and symbols represent molecular mass versus elution time (right axis). (B) Experimental SAXS profiles for Nsp7/8, Nsp8 þ dsRNA, and Nsp7/8 þ dsRNA collected at the SEC peak shown together with calculated SAXS profiles from best fitting atomic models (black line) or alternative model (dash line). Guinier plots for experimental SAXS curves are shown in the inset. Residuals of best-fit models (colored as indicated), alternative models (gray), and goodness-of-fit values (c2) are shown in bottom plot. (C) Solution model of Nsp8-dsRNA (magenta and RNA in red) used in the calculate SAXS profile in (B) with overlaid SAXS-based shape. (D) Nsp8 EMSA with radio-labeled polynucleotides shows no binding of ssDNA (right) and binding of all ssRNA substrates. (E) Ensemble of structures that fit Nsp7/8 used in the calculated SAXS profile in (B). Mass of each model is indicated. (F) The ensemble that fits the SAXS from Nsp7/8 þ dsRNA with mass of each model indicated. (G) SANS data for the Nsp7/8/RNA complex (pink circles) and Nsp7/8/DNA in 65% D_2O (light blue circles) were fit by the models shown in (E) and (C).

237

As a monomer Nsp8 adopts a different conformation to that observed in atomically resolved complexes. The conformations of Nsp8 monomers found in the Nsp7/8 crystal structure (PDBID: 3UB0,[473] and Nsp7/8/12 cryo-EM structure (PDBID: 6YYT, [471] did not match the SEC-SAXS data from the elution region containing the monomers (χ^2 =40, Figure 7-36B). In these Nsp7/8 or Nsp7/8/12/RNA complex structures, the Nsp8 N-terminal helix-turn-helix bundle (1-100) adopts various extended conformations. To find a conformation that fits the SAXS data, we employed conformational sampling of the N-terminal helix-bundle. The flexible tethers in between the head and two distinct helix regions (1-82 and 86-100) were identified by structural comparison of the two Nsp8 conformers taken from the Nsp7/8/12/RNA complex [471] (Figure 7-36D). Despite providing a nearly exhaustive search of extended and compact conformations to select an ensemble, an excellent fit to the SAXS data (Figure 7-36B) required a compact configuration with the head domain in close proximity to the N-terminal region of the helix-turn-helix region (Figure 7-36D, Figure 7-43). The folding back of the structure may be driven by a polar interaction between a positively charged segment of the long connecting helix with a negatively charged portion of the beta-barrel containing domain (head), as calculated by PDB2PQR/APBS [474,475] (Figure 7-44). These regions could be further driving the multimerization, as shown in the crystal structure of the Nsp7/8 complex of Feline Coronavirus (PDBID: 3UB0) (Figure 7-43C) [473].

To fit higher concentrations of Nsp8, we created a model for dimers that combines our monomeric model with dimerization contacts of Nsp8 found in an analogues crystal structure [473]. A mixture of the dimer and monomer models fit the SAXS data for Nsp8 at a 2.5 mg/ml concentration (Figure 7-43A) and the SAXS data from the leading edge of the SEC major elution peak. At our higher concentration (5 and 10mg/mL), the SAXS profile could not be fit using monomer/dimer mixture and required model with longer maximal distances, as shown by Dmax values determined from the P(r) function (Figure 7-43B). A small contribution of a larger tetrameric state built based on Nsp7/8 structure (PDBID: 2AHM)[476], was necessary to fit SAXS data for the two highest protein concentrations. (Figure 7-43C).

7.5.2.5   Nsp12

Nsp12 alone expresses poorly and has low solubility. This suggests that Nsp12 is unlikely to function alone and must form heteromeric complexes for stability. The maximum concentrations obtained were 2.5mg/mL, above which aggregation of the protein was evident. The MALS and SAXS mass is consistent with a monomer (Figure 7-35, Figure 7-36A). The interpretable SAXS signals below q < 0.2Å^{-1} were in good agreement (χ^2 =1.4, Figure 7-36B) with the available atomic model of Nsp12 (Figure 7-36E) [471]. The Rg of the measured profile was 31.1±0.5 Å, while that calculated from the model with added missing N-terminal region (1-31 and 50-77) is 31.6Å (Table 7-8).
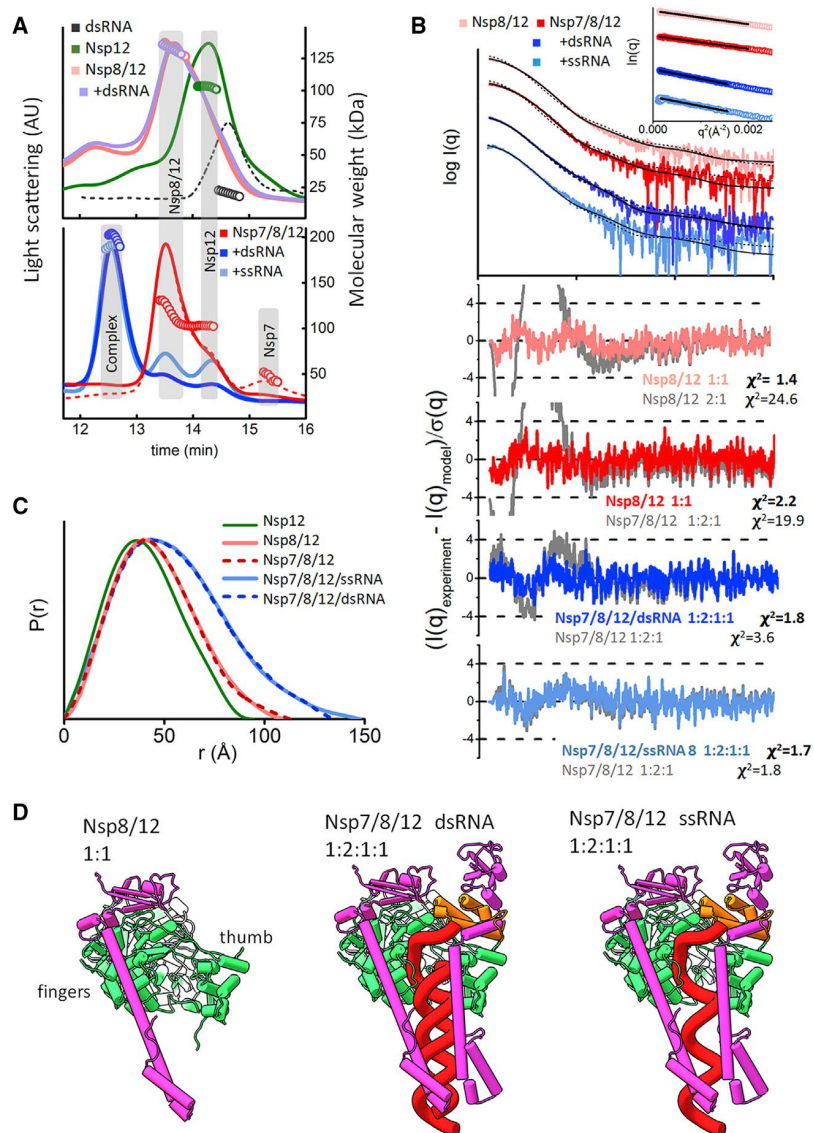
Figure 7-38. RNA-stabilized Nsp7/8/12 complex. (A) SEC-MALS chromatograms for Nsp12, Nsp8/ 12, Nsp8/12 þ dsRNA, and dsRNA (top) and Nsp7/ 8/12, Nsp7/8/12 þ dsRNA, and Nsp7/8/12 þ ssRNA (bottom) are colored as indicated. Solid lines represent the light scattering detector units, and symbols represent molecular mass versus elution time. (B) Experimental SAXS profiles for Nsp8/12, Nsp7/8/ 12, Nsp8/12 þ dsRNA, and Nsp7/8/12 þ ssRNA collected at the SEC peak are shown together with the theoretical SAXS profiles for best fitting models (black line) and alternative models (dash line). SAXS fits are shown together with the fit residuals for the solution-state model (colored as indicated), alternative model (gray), and goodness-of-fit values (c2). Guinier plots for experimental SAXS curves are shown in the inset. (C) Normalized P(r) function for Nsp12, Nsp8/12, Nsp7/8/12, Nsp7/8/12 þ dsRNA, and Nsp7/8/12 þ ssRNA. The similarity of P(r) functions between Nsp8/12 and Nsp7/8/12 further confirms the absence of Nsp7 and one Nsp8 in the Nsp7/8/12 mixture. (D) Solution-state models for Nsp8/12, Nsp7/8/12 þ dsRNA, and Nsp7/8/ 12 þ ssRNA were used to fit experimental data shown in (B).

7.5.2.6   Nsp8 binds RNA but not DNA.

We characterized the interaction of Nsp8 with dsRNA (Figure 7-37, Figure 7-42) and dsDNA (Figure 7-44). The dsDNA is analogous in sequence to dsRNA. Nsp8 did not form a complex with dsDNA as both the Nsp8 and dsDNA elute separately at elution times consistent with each run independently. In contrast, Nsp8 forms two types of complexes with the dsRNA contributing to two well-separated peaks (Figure 7-37A). The first elution peak (13.5 min after injection) is consistent with a 1:1 Nsp8 dsRNA complex while in the second Nsp8/dsRNA peak (14.3 min) the mass decreases rapidly across the elution profile, indicating the presence of a heterogeneous mixture of Nsp8 and free dsRNA.

We further tested whether Nsp8 binding to DNA/RNA is length or sequence dependent. Nsp8 did not bind either long (36 base) or short (28 base) ssDNA in our SEC-MALS-SAXS analysis nor in an electrophoretic mobility shift assay (EMSA) (Figure 7-37D). Nsp8 did bind the long ssRNA but not the short ssRNA at high affinity. The mixture of Nsp8 with long ssRNA shows only one distinct peak consistent with 1:1 complex as judged by determining the molecular mass and SAXS parameters (Figure 7-35, Figure 7-37A). We further tested the binding of long ssRNA, short ssRNA and the long ssRNA without the poly-U 5' end by EMSA. Nsp8 showed evidence of binding all three, however affinities are difficult to assess with EMSA bands. The EMSA bands with ssRNA are split. On the same mixtures from SEC-MALS-SAXS, the mass through the peak drops rapidly (Figure 7-37A) suggestive that multiple Nsp8s could bind the same ssRNA. These results demonstrate Nsp8 discriminates between RNA and DNA. SEC-MALS-SAXS showed a preference for longer ssRNA over the shorter.

We found Nsp8/dsRNA models that fit the SAXS data from the highest quality signal in the constant mass region of the main elution peak (Figure 7-37A, B). This model is supported from the elongated shape determined from SAXS by DAMMIF [472], and consists of Nsp8 binding dsRNA along their long axes (Figure 7-37C). In our rigid body modeling using BILBOMD [283] we preserved the Nsp8-dsRNA interaction at the N-terminal Nsp8 helix bundle that contains positively charge patches (Figure 7-44) and sampled conformations of C-terminal head domain. An excellent fit ($\chi^2 = 1.7$, Figure 7-37B) was found with the Nsp8 head domain wrapping around the RNA capping the terminal region of the RNA molecule (Figure 7-37C) providing a structural model for Nsp8 RNA binding.

7.5.2.7   Nsp7/8 complex is transient in solution.

The shape of the elution peak from co-expressed and affinity tag purified Nsp7/8 is consistent with a transient complex composed of oligomers and free proteins. The SEC-MALS-SAXS elution is dominated by an asymmetric peak with a long tailing shoulder region (Figure 7-37A, lower). The mass at the leading edge by MALS is 45kDa and SAXS is ~42kDa, which is smaller than the tetramer's expected molecular mass (63kDa) (Figure 7-35). Furthermore, a well-separated peak observed at later elution times is consistent with that observed from Nsp7 dimers, further supporting a transient Nsp7/8 complex (Figure 7-37A).

To fit the SAXS data from the dominant peak, we used an ensemble of models. The ensemble pool included a dimer and hetero-tetramer of Nsp7/8 built from our crystal structure where Nsp8 is in an extended conformation to include a helix and N-terminal bundle. We also included a

monomeric Nsp8 with various conformations of the N-terminal helix bundle, and an Nsp8 dimer with open and compact head domain (see Figure 7-43C). The data is the best fit by a mixture of 26% Nsp7/8 heterotetramer, 56% monomeric Nsp8, and 18% Nsp8 dimer with a compact head domain (χ^2 = 2.2 Figure 7-37B and E). The Nsp7/8 heterotetramer observed in our crystal structures is a minor population at these concentrations.

A similar analysis was performed on SANS data collected from the same complex differing by collection at high concentration and in a fully equilibrated sample (non-SEC). The mass determined from SANS was also slightly lower (56 ± 3 kDa) than the Nsp7/8 heterotetramer. The Nsp7/8 heterotetramer crystal structure did not fit the SAXS data. However, using OLIGOMER a good fit ($\chi^2 < 1.0$) was obtained with a mixture of 88% Nsp7/8 heterotetramer and 12% Nsp8 monomer. This supports the SEC-SAXS analysis that shows the dynamic and concentration dependent nature of the interaction between Nsp7 and Nsp8.

Figure 7-39. Assembly of the RTC components. The solved crystal structure (top left) reported in this work exists in a dynamic equilibrium and forms at high concentrations of Nsp7 and Nap8. Nsp7 is dominantly dimeric on its own (bottom left), though it can form linear oligomers. Nsp8 alone is in a compact conformation (center). However, this conformation becomes extended when RNA or Nsp7 binds in a competitive manner for available binding sites on Nsp8. When all three are combined with a stabilized form of 1:1 Nsp8/12, a very stable Nsp7/8/12 RNA complex forms in a 1:2:1:1 ratio for RNA transcription. The architecture, preferred binding sites for Nsp8 on the fingers domain, and strong binding of RNA by Nsp8 suggest a mechanism in which the Nsp8 on the thumb domain may swap positions with that on the fingers while the one on the fingers progresses with the RNA.

Further, to test our findings that Nsp7 induces structural change in Nsp8 relative to its monomeric state, we performed contrast matching SANS experiments with the Nsp7/8 complex (Figure 7-36F). Deuterated Nsp7 (dNsp7) was co-purified with protiated Nsp8 to form dNsp7:Nsp8. Using the contrast match point of dNsp7 (90% D_2O) we single out the scattering from complexed Nsp8. Figure 3F compares the pair distribution function (P(r)) of dNsp7:Nsp8 against a solution of Nsp8 without Nsp7. The P(r) from the complex, dNsp7:Nsp8 has a bimodal shape indicating two well separated protein densities while the P(r) of Nsp8 alone is more centralized to one main distance. The change verifies Nsp7 dramatically modifies Nsp8 conformation or assembly.

7.5.2.8   RNA competes with Nsp7 to interact with Nsp8.

In contradiction to several proposed pathways, our SEC-MALS-SAXS experiments do not support a stable ternary Nsp7/8 dsRNA complex. Quite the opposite, our results indicate a competition between Nsp7 and dsRNA for Nsp8 binding. The elution peak from mixtures of Nsp7/8 with RNA is shifted in agreement with the smaller measured mass (MALS) from 45kDa for Nsp7/8 to ~35kDa for Nsp7/8 +dsRNA (Figure 7-37A, lower). The secondary SEC peak for free Nsp7 is more pronounced and well separated (Figure 7-37A), suggesting the uncoupling of Nsp7 from the complex.

Fitting of the SAXS data provides further evidence that dsRNA destabilized the Nsp7/8 interface. From the pool of models that include conformers of Nsp8, Nsp7 monomer and dimer, Nsp7/8 dimer and tetramer, and multiple conformers of Nsp8/dsRNA, the best fit was obtained with the Nsp8/dsRNA complex ($\chi^2 = 1.7$, Figure 7-37B) described above. The multistate model selection that contains Nsp7 dimer further improved the SAXS fit ($\chi^2 = 1.6$, Figure 7-36B,D) and suggested transient binding of dsRNA that led to the decoupling of Nsp7/8 and presence of free Nsp7 across the peak. The presence of Nsp7 dimers also explains the smaller Rg and mass determined by SAXS or MALS relatively to the values measured for the Nsp8/dsRNA complex (Rg 28.9Å vs. 32.2 Å; MW_{MALS} 32 vs. 45 kDa; MW_{SAXS} 30 vs. 45kDa) (Figure 7-37A, Figure 7-35, Table 7-8).

Contrast matching SANS experiments, masking the RNA signal in 65% D_2O, were also performed on Nsp7/8 dsRNA mixtures. The contrast matched SANS profile is best fit ($\chi^2= 1.2$) with Nsp8 monomer (Figure 7-37G), supporting the dissociation of Nsp7/8 complex observed in the SEC-SAXS. Altogether, the combined SANS and SAXS indicate that RNA alters the interactions found in our crystal structure of Nsp7/8 alone leading to the formation of a smaller Nsp8/dsRNA complex and disassociation of Nsp7 from the complex (Figure 7-37F).

7.5.2.9   Without Nsp7, Nsp12 recruits one Nsp8 to its finger region and does not bind RNA.

The mass across the center of the SEC-MALS-SAXS main peak from co-expressed Nsp8 and 12 is 130 kDa, in agreement with the mass of an Nsp8/12 1:1 complex (Figure 7-38A). The main peak trails into a mass (110kDa) consistent with a free Nsp12. The broadening of the elution peak indicates disassociation of Nsp8 from Nsp12. The central portion of the main peak was merged to obtain a SAXS profile sufficient to be modeled in detail.

The Nsp8/12 SAXS profile was fit with a pool of the models containing various conformers of Nsp8 (Figure 7-36D), Nsp12 (Figure 7-36E), and three models of the Nsp8/12 complex with Nsp8 located at the fingers, thumb region, or both [471]. The pool also contained alternative models of the Nsp8/12 complex with an extended Nsp8-helix bundle region. The best fit (χ^2 =1.4) was obtained with a single model of Nsp8/12 containing one Nsp8 bound to the finger region (Figure 7-38D). This model matches experimental SAXS better than the Nsp8/12 model with both Nsp8 bound (χ^2 = 24.6) (Figure 7-38B) and agree with the determined mass (Figure 7-35). We attempted to improve the fit with the multistate models, but the single conformation remained the best fit.

These findings suggest Nsp12 recruits only one Nsp8 to the finger region and raise the question of whether Nsp7, RNA, or both are required to stabilize the Nsp8 interaction at the thumb region. Therefore, we investigated Nsp7/8/12 assembly in solution and the interactions of Nsp8/12 with RNAs.

Nsp7, Nsp8, and Nsp12 were co-expressed and purified for SEC-MALS-SAXS analysis. The first SEC trace, prior to SAXS measurements, shows a split peak and tail with a significant peak from Nsp7 at later elution times, already suggesting transient dissociation of Nsp7 from the Nsp8/12 complex (Figure 7-46). The early fractions were subsequently analyzed (and consequently purified by SEC a second time) by SEC-MALS-SAXS. The mass of the first elution peak (132kDa) agrees with the Nsp8/12 1:1 complex. The elution time of the peak (Figure 7-38A), determined mass, SAXS parameters (Figure 7-35), and calculated P(r) functions (Figure 7-38C) from this first peak are identical to those reported above from Nsp8/12 when no Nsp7 was present. Furthermore, the Nsp8/12 1:1 model (Figure 7-37D) gives an excellent match to the experimental SAXS curve (χ^2 =2.2) (Figure 7-38B) that is distinct from other potential models, including an Nsp7/8/12 1:2:1 complex with the significantly worse fit (χ^2 =19.9). The trailing shoulder and peak from the elution trace is consistent with Nsp12 alone and an Nsp7 dimer (Figure 7-38A), further supporting the dissociation of Nsp7 from the Nsp8/12 1:1 complex.

Mixing an excess of RNA with Nsp8/12 does not lead to a high-affinity complex. Identical elution time and MALS-measured molecular weights (~130kDa) of the early elution peak of Nsp8/12 +dsRNA (Figure 7-38A), Nsp8/12 + ssRNA (Figure 7-46), and Nsp8/12 alone clearly shows the absence of an Nsp8/12 interaction with RNA. The low or non-existent binding of RNA by Nsp8/12 is surprising since Nsp8 alone binds RNA strongly. The finding that Nsp8/12 does not stably interact with Nsp7 or RNA on its own, Nsp7 needs to interact with Nsp8 to form complete SARS Cov-2 polymerase machinery.

7.5.2.10 Nsp7/8/12 requires RNA for stability.

Adding dsRNA or ssRNA stabilizes the Nsp7/8/12/RNA within a 1:2:1:1 complex as visualized in cryo-EM atomic resolution structures [471]. Thoroughly mixing near equivalent molar ratios of dsRNA or ssRNA with the initially purified Nsp7/8/12 complex assuming a 1:2:1 complex yields a sharp and near symmetric SEC elution peak (Figure 7-38A and Figure 7-45). The shift in the SEC peak and determined mass (~190kDa) suggest the formation of Nsp7/8/12/RNA in 1:2:1:1 complex for both complexes, that agrees with the theoretical mass of 1:2:1:1 complexes with 181

kDa for Nsp7/8/12/ssRNA and 172kDa Nsp7/8/12/dsRNA (Figure 7-35). The experimental SAXS profiles (Figure 7-38B) for both complexes are consistent with models built from cryo-EM structure [471] by adding all missing regions in the protein and RNA region and removing a shorter RNA strand for Nsp7/8/12/ssRNA complex (Figure 7-38D). This agreement further suggests that both Nsp8 copies remain pointed in the direction of RNA polymerization rather than flexing in alternate conformations.

7.5.3   Discussion

Our overall aim was to further develop a comprehensive understanding of the SARS-CoV-2 RTC and identify strategies for inhibiting its assembly and mechanism. To this end we have integrated information from available cryo-EM and X-ray crystallography structures with results reported above, which include, biochemical assays, atomic resolution structures of Nsp7/8, and 20 SAXS and SANS experiments on components of the RTC. Our crystal structures, those resolved by others and cryo-EM results provide a basis to interpret our solution scattering studies that probe the transient and plastic properties of the assembly which are difficult to attain with static structures alone. The highly evolved and dynamic nature of this essential complex greatly enables SARS-CoV-2; however, these same properties also provide many avenues for disruption and inhibition. Summarizing our results, we have arrived at the following insights.

At high concentrations, as found in our crystal structure, Nsp7/8 forms a tetrameric structure composed of a dimer of heterodimers. However, this structure is transient and at low concentrations the proteins disassociate into mixtures of the individual components and smaller complexes. Nsp7 primarily exists as a dimer, but longer oligomeric structures are also present which can be confounding to size-based purification methods since Nsp7 will be present in one oligomeric form or another at separate elution times. Without being bound to Nsp7, Nsp8 adopts a compact conformation not yet observed in crystal structures and will oligomerize into flat aggregates at high concentrations. Nsp8 will readily bind dsRNA and ssRNA but not ssDNA. Nsp12 alone is unstable and prone to aggregation even at low concentrations (>2.5 mg/ml). Binding of Nsp8 markedly improves stability.

The atomic coordinates from our crystal structure and our solution scattering results inform on the consequences of Nsp7's overall hydrophobic character. The eight hydrophobic amino acids make up 45% of the protein including 18% Leucine and 11% Valine. The solvent excluded interface made between Nsp7 and Nsp8 is 57% hydrophobic. In all of our structural studies, Nsp7 is found in complex with Nsp8 and Nsp12 or it has the propensity to form chains of NSP7 oligomers on its own. A further consequence of its hydrophobic nature is its ability to disrupt the hydrophobic cores of the other proteins and modify their conformational states as exemplified by its interaction with Nsp8. The 2:2 Nsp7/8 we observe in our crystal structure may be a storage form of the complex when little RNA is available. Nsp7's ability to compete with RNA, a hydrophilic and charged molecule, for Nsp8 binding is not intuitive. One mechanistic possibility is that the binding of Nsp7 reduces the footprint Nsp8 has to bind RNA and serves to offload RNA from Nsp8. This mechanism could be fundamental to the operation or assembly of the RTC and provides a rational for expressing Nsp7 as a separate peptide rather than as part of Nsp8 or Nsp12.

We observed two stabilized structures with Nsp12. In the absence of its binding partners, Nsp12 alone seems inherently unstable which limited our ability to create complexes by mixing the purified proteins. However, by co-expressing the individual proteins, reasonable yields of the complexes could be attained. A stable 1:1 structure form between Nsp8 and Nsp12, where an Nsp8 sits on the fingers of Nsp12. This complex, however, is not capable of binding the RNAs we tested. Although cryo-EM has partially visualized an Nsp7/8/12 structure without RNA, the sample had to be supplemented with significant concentrations of Nsp7 and 8 for resolving these structures. In our hands and at the concentrations we made our measurements (1-5mg/mL), Nsp7/8/12 does not form a stable structure until RNA is present. Once RNA is present, a very stable 1:2:1:1 Nsp7/8/12/RNA complex is formed both with dsRNA and ssRNA.

In all the most complete cryo-EM structures of the Nsp7/8/12/RNA with and without Nsp13 or Nsp9, the structure and position of Nsp7 and 8 are remarkably similar. This similarity suggests Nsp7 and 8 are static components during replication. Their role is thought to be three-fold, to help close the complex once RNA is bound, guide RNA upon exit and stabilize contacts with Nsp13 and other proteins. As noted with the first complete structure, Nsp7 and 8 confer processivity [471]. The orientations of the helical extensions on Nsp8 remain uniquely observed in the RTC and may function similar to sliding clamps in DNA replication. Nsp8's structure is suited to preforming or further enhancing a straightened RNA with positively charged RNA binding patches spaced to coincide with an RNA double helix.

However, for these roles, expression of Nsp7 and Nsp8 as independent polypeptides, rather than as one protein or as part of Nsp12, seems inefficient. Nsp7 and Nsp8 have been postulated to take part in other viral activities which may justify their independence. Exploring other potential avenues, we build upon cryo-EM and our results to suggest more dynamic roles for the components. We propose that Nsp8 is involved in recognizing ssRNA over ssDNA and guiding ssRNA to or from the RTC complex. Our observation that Nsp8 adopts a different conformation alone relative to bound states suggests Nsp8 is a dynamic component of the macromolecular machine. Based on the exclusive binding of either Nsp7 or RNA by Nsp8 but not both simultaneously, we propose that Nsp7 promotes the release of RNA from Nsp8 once it has been guided to the RTC by disrupting Nsp8's head domain. The stability of the Nsp8 on the fingers domain is intriguing and certainly serves to further stabilize the RTC's RNA binding. However, we have shown that the fingers-bound Nsp8 is not sufficient to bind RNA and Nsp7 is necessary.

The binding and position of two Nsp8s could play a role in transitions that occur during RNA transcription. RNA rotates and progresses during polymerization. The compacting and expanding properties of Nsp8 could be part of a retraction mechanism where, once the N-terminus of Nsp8 is overstretched, it releases and re-binds closer to Nsp12 where newly synthesized and duplexed RNA is emerging. Alternatively, the combination of RNA rotation and progression may pull the fingers-bound Nsp8 off of Nsp12 and its vacant spot may be taken up by the Nsp8 bound to the thumb (Fig 6). The energetics could be favorable since there is a 1:1 trade from an Nsp8 in a slightly destabilized configuration, through rotation and progression of the RNA. The now vacant spot on the thumb domain would then be occupied by either free Nsp8 or perhaps both Nsp7 and 8 are replaced. The transition would also allow Nsp12 to reset or release its grasp on RNA since binding of RNA by Nsp12 is dependent on Nsp7. The available structures of Nsp7/8/12/13/RNA do not prohibit such an exchange and rather suggest that Nsp13 could also be

245

part of the transition with the template strand eventually released by the Nsp13 helicase and grabbed by a new one bound at the thumb as well.

The possible mechanistic roles of the components in the RTC described above (Figure 7-39), require significant further investigation. In addition to the continued clarity provided by high resolution cryo-EM and crystallographic structures, further solution and cell-based studies, to examine the dynamics of RTC in its many contexts, will be helpful. For example, RNA transcription and polymerization appear to occur in double membrane vesicles where the concentrations of the components may be manipulated. We observed several concentration dependent assemblies in our studies. Additionally, this study was conducted with a limited and specific set of RNAs that may bias our findings. Further studies with a variety of RNA substrates are necessary.

7.5.4   Supplementary Material

Table 7-7. Data Collection and Refinement Statistics.

	Nsp7-8A	Nsp7-8B	Nsp7-8C
Crystallization Conditions			
Data Collection			
Space group	$C222_1$	$P2_1$	$P2_1$
Unit cell parameters (Å; º)	$a = 52.15, b = 70.78, c = 115.46; \alpha = \beta = \gamma = 90.0,$	$a = 47.26, b = 55.10, c = 84.40; \gamma = 90.0, \beta = 90.4$	$\alpha = a = 35.62, b = 119.90, c = 43.38; \alpha = \gamma = 90.0, \beta = 92.2$
Resolution range (Å) [a]	50.00 – 2.85 (2.90 - 2.85)	50.00 - 1.95 (1.98 - 1.95)	50.00 – 1.50 (1.53 - 1.50)
No. of reflections	5,144 (235) [a]	30,923 (1,220)	56,701 (2,392)
R_{merge} [b]	0.13 (1.11)	0.18 (1.15)	0.083 (0.84)
Completeness (%)	97.9 (88.7)	97.3 (77.9)	97.7 (83.9)
$CC_{1/2}$ [c]	0.853 (0.468)	0.962 (0.456)	0.999 (0.826)
$I/\sigma(I)$	18.2 (1.01)	22.5 (1.1)	30.2 (1.85)
Multiplicity	6.7 (4.4)	8.7 (2.2)	4.7 (3.0)
Wilson B factor	105.9	38.6	22.8
Structure Determination			
MR initial model (PDB ID)	5F22	6WIQ	6WIQ
Refinement			
Resolution range (Å)	35.39 - 2.85 (3.59 - 2.85)	41.12 - 2.00 (2.01 - 1.95)	40.76 - 1.50 (1.53 - 1.50)
Completeness (%)	97.2 (96.0)	96.2 (73.0)	97.6 (82.0)
No. of reflections	5,108 (2,462)	30,700 (2,095)	56,635 (2,468)
R_{work}/R_{free} [d] (%)	20.8/25.2 (31.2/32.3)	18.7/22.9 (32.6/41.4)	16.1/19.9 (21.6/25.5)
Protein chains/atoms	2/1,507	4/3,077	4/2,988
Ligand/Solvent atoms	-	12/84	20/235
Mean temperature factor (Å²)	125.6	53.1	35.5
Coordinate Deviations			
R.m.s.d. bonds (Å)	0.001	0.008	0.011
R.m.s.d. angles (º)	0.357	0.829	1.158
Ramachandran plot [e]			
Favored (%)	96.86	98.04	99.19
Allowed (%)	2.62	1.68	0.81
Outside allowed (%)	0.52	0.28	0.0
PDB Accession Code	6WIQ	6WQD	6XIP

[a] Values in parentheses correspond to the highest resolution shell.

[b] Rmerge = ΣhΣj|Ihj–<Ih>|/ΣhΣjIhj, where Ihj is the intensity of observation j of reflection h.

[c] As defined by Karplus and Diederichs [477]. [d]R = Σh|Fo|–|Fc|/Σh|Fo| for all reflections, where Fo and Fc are observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement. [e]As defined by Molprobity [478]

Table 7-8. Structural parameters from SAXS, SANS and MALS data.

SAXS sample SASBDB#	D_{max} (Å)	Rg (Å) from Guinier plot	Rg (Å) from P(r)	MW Seq. monomer (kDA)	MW SAXS (kDa)	MW MALS (kDa)	Model fit χ^2	data source
Nsp7 JUFZRI	~ 65	18.6± 0.6	18.8	9	19	17	1.5	SEC-SAXS peak
Nsp8 IDDWOG	~ 90	24.6±0.7	25.8	22	29	28	1.4	SEC-SAXS tail
Nsp8 G1TQHV	~ 110	26.8±0.1	27.8	22	33	35	1.9	SEC-SAXS peak
Nsp8	~ 110	29.9± 0.9	30.6	22	38	ND	ND	SANS
Nsp12 QIWPPC	~ 110	31.1± 0.5	41.9	107	88	93	1.3	SEC-SAXS
Nsp7/8 DJTRUW	~ 100	28.8±0.1	30.3	31	42	45	2.2	SEC-SAXS peak
Nsp7/8	~ 110	30.0± 0.1	ND	31	56	ND	ND	SANS
Nsp8+dsRNA 7MRNJA	~ 120	32.2± 0.6	29.8	45	46	45	1.7	SEC-SAXS peak
Nsp8+ssRNA XMU3AK	~105	30.3± 0.2	31.4	33	35	35	ND	SEC-SAXS peak
Nsp7/8 +dsRNA TKZWCR	~ 105	28.9± 0.2	31	54	32	30	2.1	SEC-SAXS peak
Nsp8/12 PPENAB	~ 115	34.8± 0.5	35.2	129	118	131	1.4	SEC-SAXS peak
Nsp7/8/12 X1TKUI	~ 115	34.7± 0.7	35.2	138	114	132	2.2	SEC-SAXS peak
Nsp7/8/12 +dsRNA ZWMCYC	~ 140	41.8± 1.7	43	159	170	195	1.8	SEC-SAXS peak
Nsp7/8/12 +ssRNA E38FZM	~ 145	41.1± 2.2	43	126	170	190	1.7	SEC-SAXS peak

a) Maximal dimension (Dmax) defined from pair distribution function (P(r)) calculated by SCATTER

b) Theoretical MW calculated for monomer or assembly within the equal molar ratio.

c) MW determined by volume of correlation Vc [334]

d) Goodness of fit (χ^2) for the atomistic models is determined FOXS[291,336] for single model or MultiFOXS[479] for multistate model.
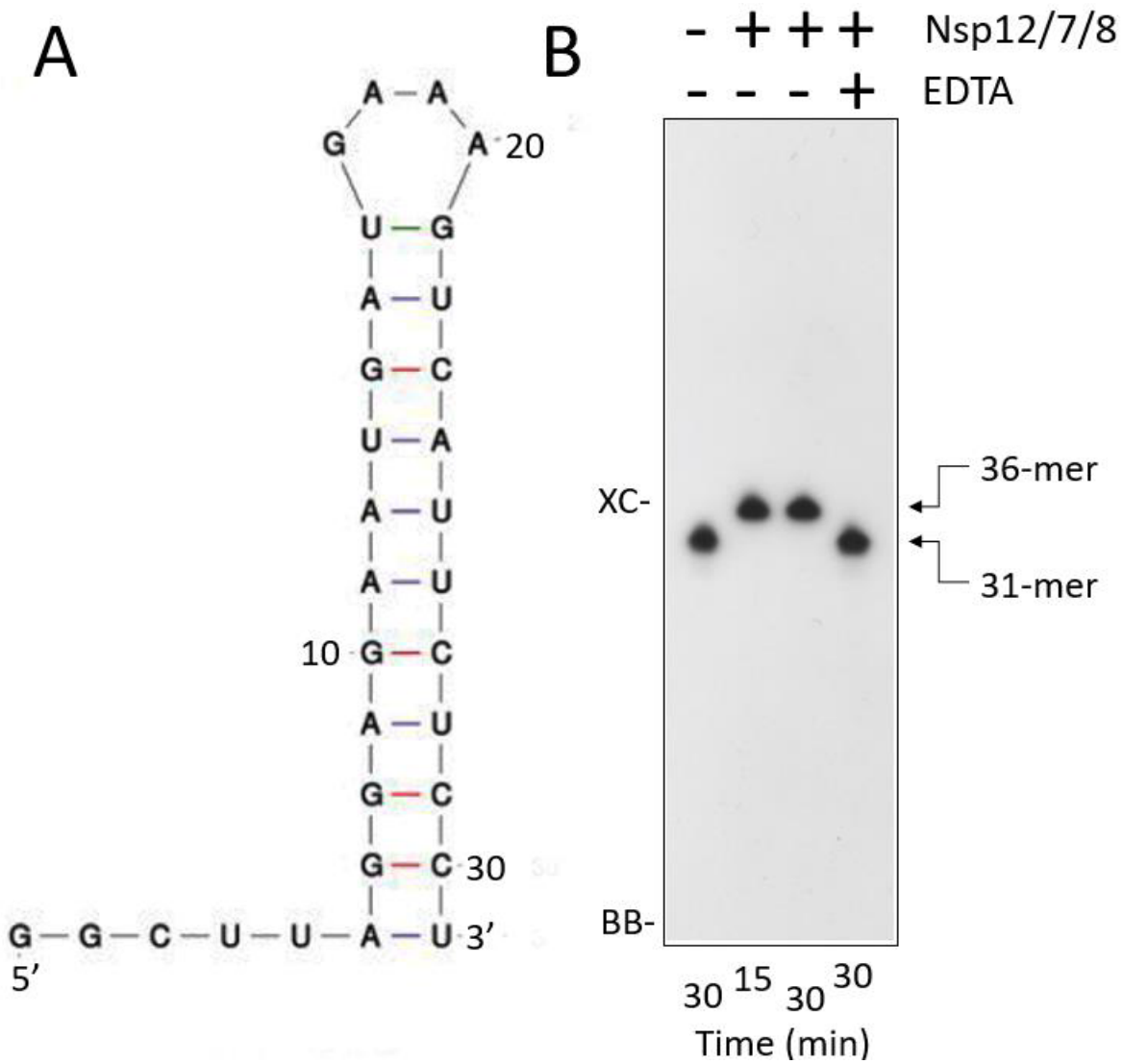
Figure 7-40. SARS-CoV-19 RTC activity. (A) Minimal substrate for RTC: a hairpin composed of 31 nucleotides. (B) Analysis of the minimal substrate extension by the Nsp12/7/8 complex on a 15% denaturing polyacrylamide gel. The hairpin was [32P]-labeled and its extension was monitored for 30 min. 2.5 mM EDTA inhibits the extension reaction. BB and XC denote positions bromophenol blue.
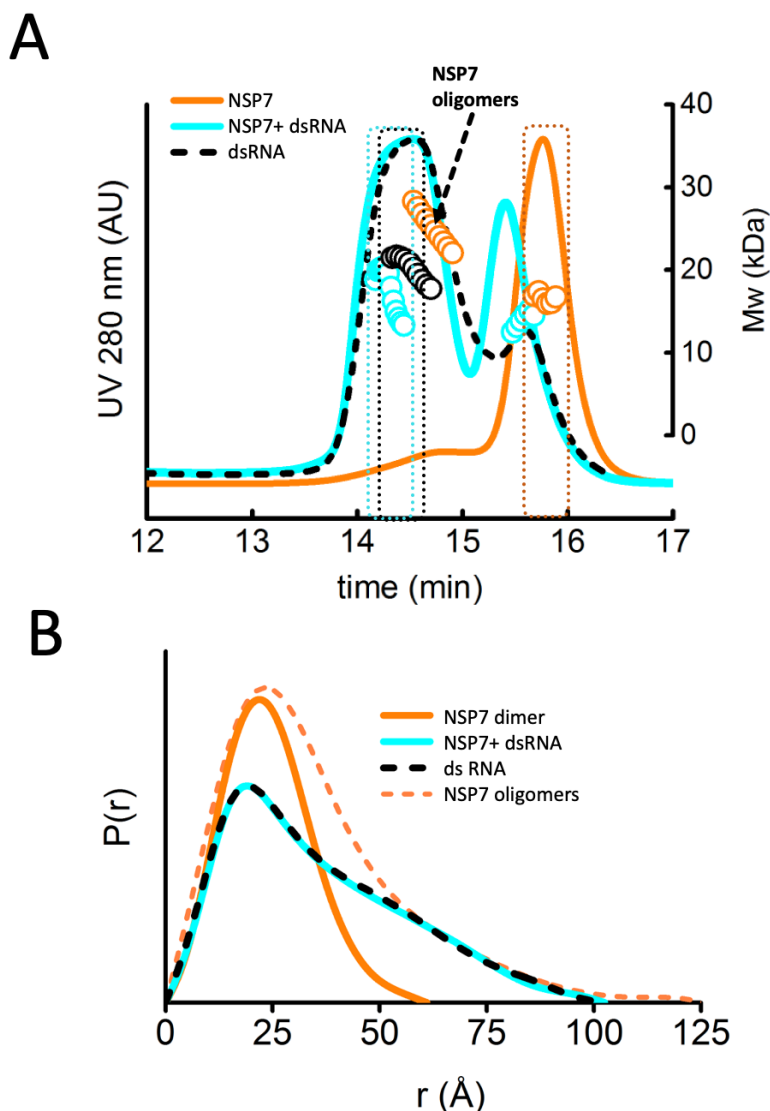
Figure 7-41. Nsp7 does not interact with dsRNA. (*A*) SEC-MALS chromatograms for Nsp7, Nsp7+dsRNA, and dsRNA are colored as indicated. Solid lines represent the UV at 280nm in arbitrary units, while symbols represent molecular mass versus elution time. (*B*) Normalized P(r) function calculated for SAXS profiles of Nsp7 dimer, Nsp7+dsRNA, and dsRNA collected at the SEC peak and Nsp7 oligomers collected at the SEC-leading peak. The similarity of P(r) functions between Nsp7+dsRNA and dsRNA further confirm the absence of Nsp7+dsRNA complexation.
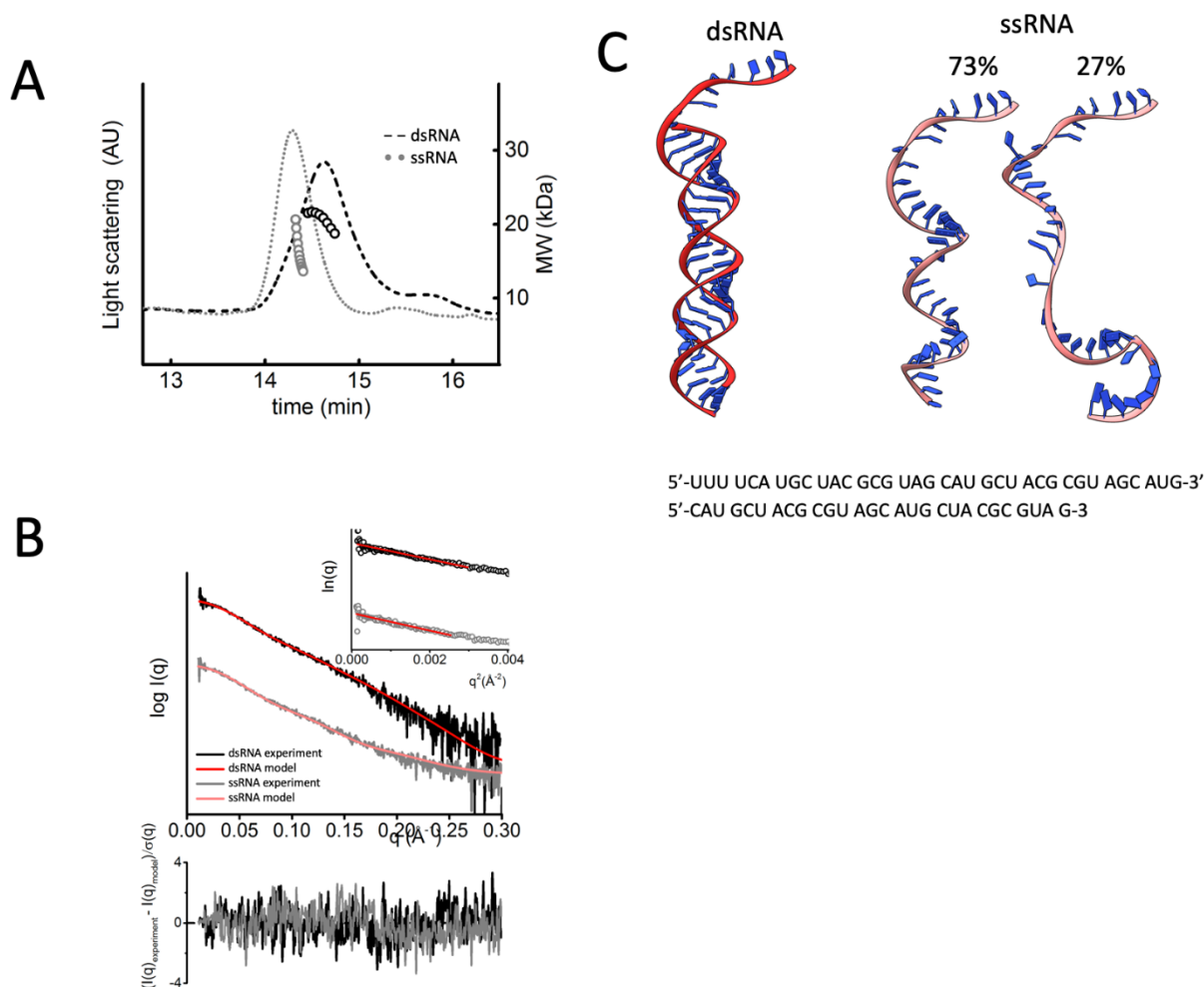
Figure 7-42. Monodisperse dsRNA and ssRNA. (*A*) SEC-MALS chromatograms for dsRNA and dsRNA are colored as indicated. Solid lines represent the UV at 280nm in arbitrary units, while symbols represent molecular mass versus elution time. (*B*) Experimental SAXS profiles for dsRNA (black) and ssRNA (gray) collected at the SEC peak are shown together with the theoretical SAXS profiles for solution-state models of ds RNA (red) and ssRNA (light red) shown in panel C. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in the inset. (*C*) Solution state models for dsRNA and ssRNA were used to fit experimental data shown in panel B.
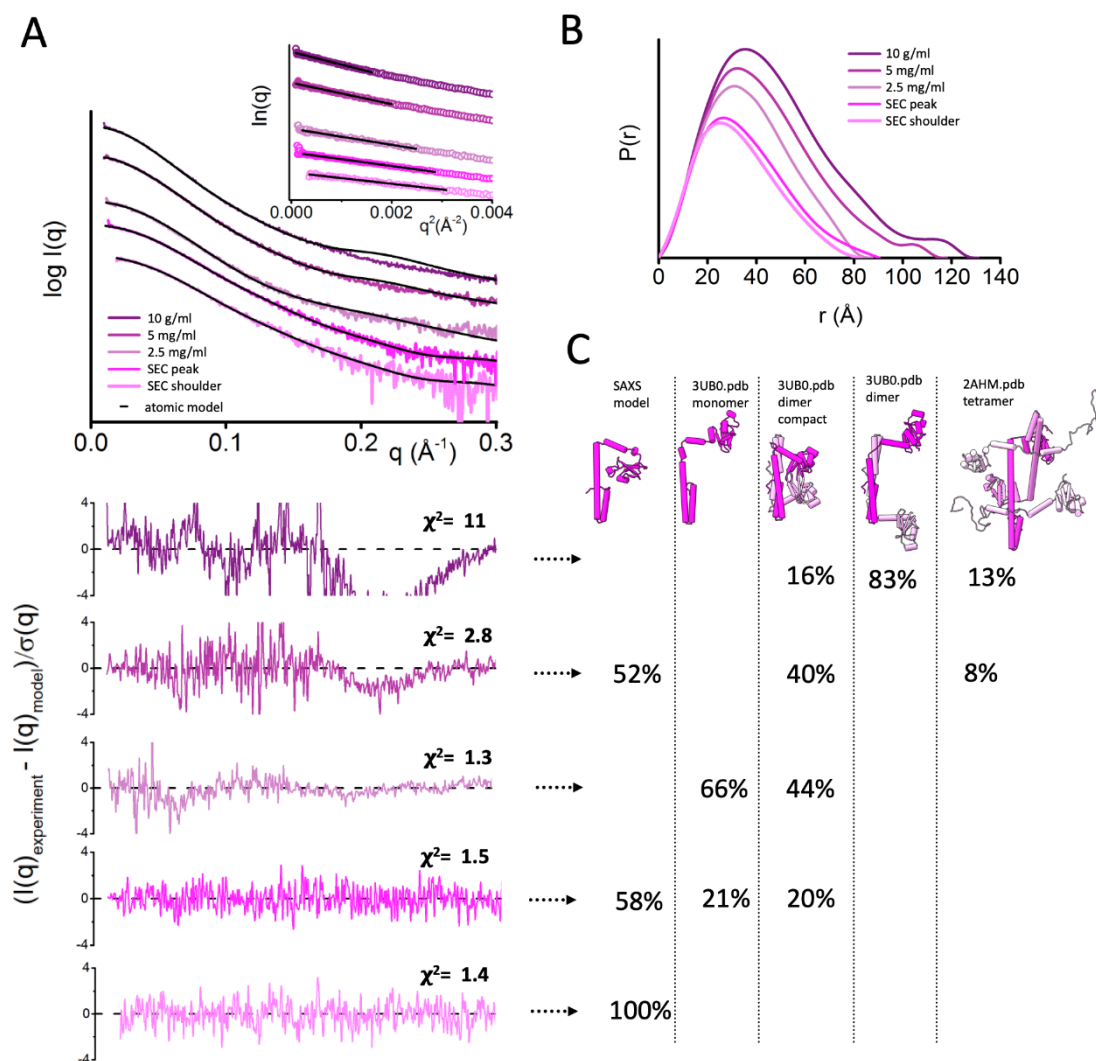
Figure 7-43. Nsp8 forms a dimer and larger oligomer at high concentration. (*A*) Experimental SAXS profiles for Nsp8 monomer, dimer, and the sizeable oligomeric state collected at SEC-SAXS elution shoulder and peak compared to the SAXS data collected in HT-SAXS mode at 2.5, 5, and 10 mg/ml (colored as indicated). SAXS data are shown together with a theoretical SAXS profile for multistate models shown in panel C and weights as indicated. SAXS fits are shown together with the fit residuals and goodness of fit values (χ^2). Guinier plots for experimental SAXS curves are shown in the inset. (*B*) P(r) functions calculated for the experimental SAXS profiles from panel A are normalized on the molecular mass determined by the volume of correlation Vc [334] (*C*) Solution state models for Nsp8 at various concentrations that fit experimental data shown in panel A.
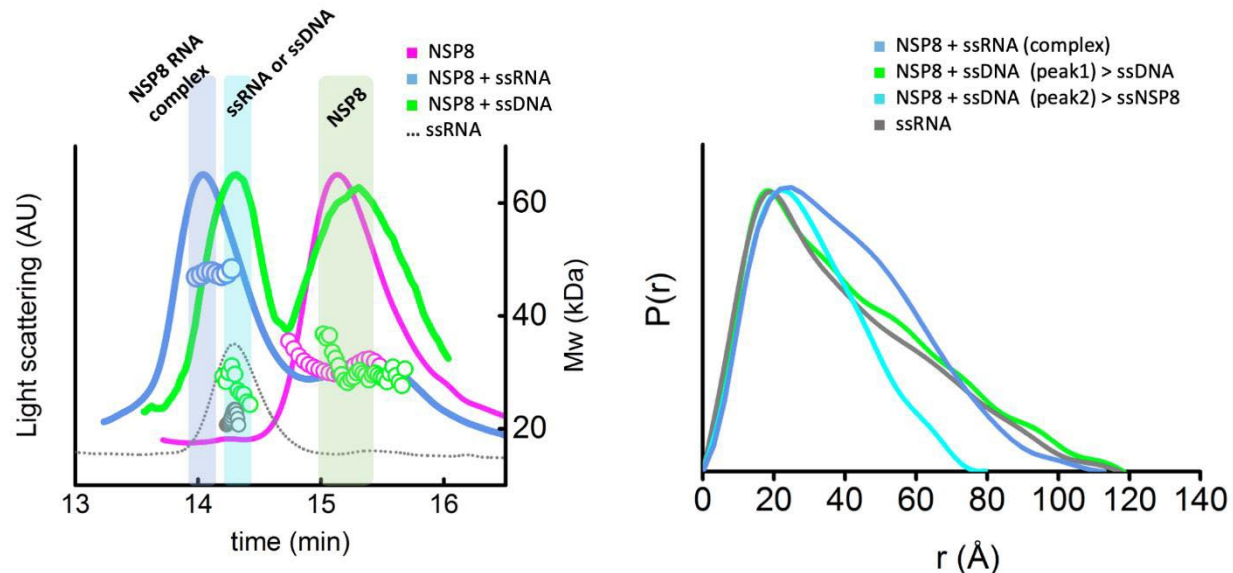
Figure 7-44. Nsp8 binds ssRNA but not ssDNA. (left) SEC-MALS-SAXS elution profiles of Nsp8, ssRNA, Nsp8 with ssRNA, and Nsp8 with ssDNA. The ssRNA is an analogous sequence and has the same length as the ssDNA. An analysis of the MALS data in each peak provides a mass as indicated by the circular points with values indicated on the right axis. The SAXS curves that were extracted from each peak containing either ssDNA or ssRNA are shown (left). The Nsp8 + ssRNA is distinct from all other curves.
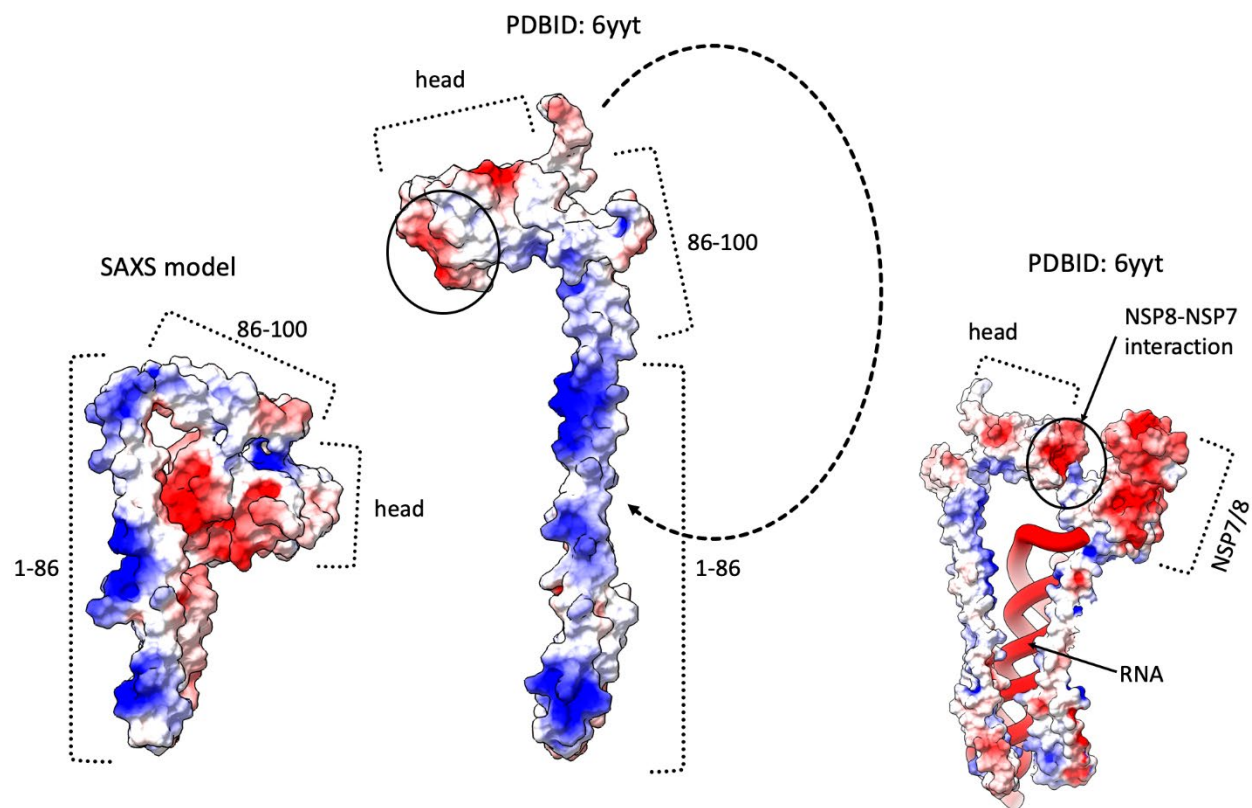
Figure 7-45. Distribution of electrostatic surface potential and flexibility of the Nsp8 helix bundle drive the compaction of Nsp8. Electrostatic surface potential for Nsp8 for SAXS model and Nsp8 conformer taken from the cryo-EM structure of Nsp7/8/12/dsRNA (PDBID: 6yyt) calculated at the pH= 7.0. The surface potential indicates that the flexible N-terminal helix bundle region (1-100) is mostly positively charged and is suited for nucleic acid-binding. In the absence of Nsp7 or Nsp12 (right panel), the negatively charged head region folds back (SAXS model) and interacts with the positively charged helix-bundle region.
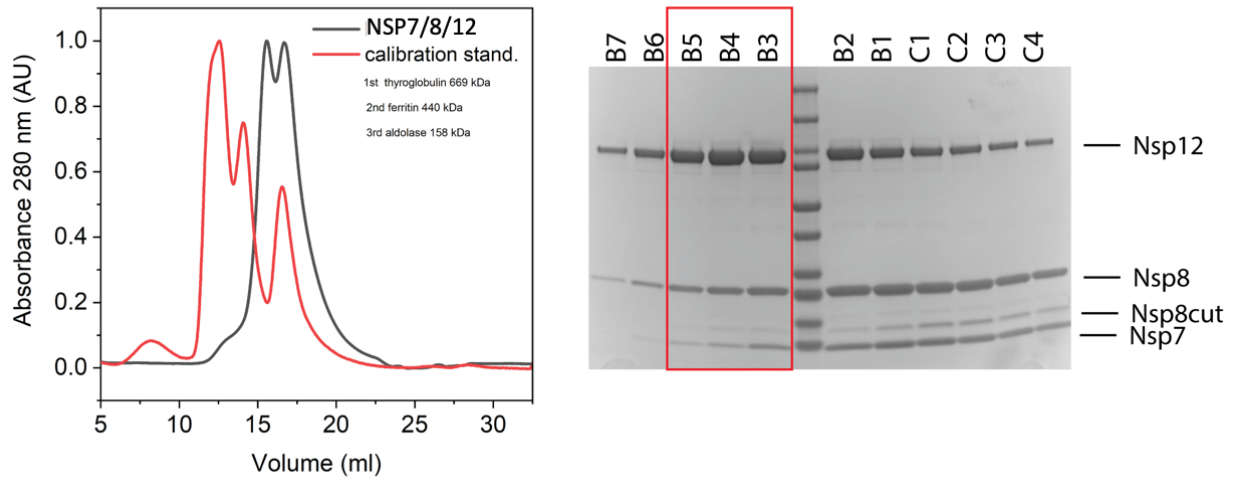
Figure 7-46. Initial SEC purification of Nsp7/8/12 sample. Splitting of SEC peak (left panel) shows heterogeneity of Nsp7/8/12 complex. The shift of Nsp8, Nsp7 bands relative to the Nsp12 on the SDS-PAGE gel (right panel) further indicates weak Nsp7/8/12 complex.

7.5.5 *Materials and Methods*

7.5.5.1 Protein Sequences

Nsp7:SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSL
LSVLLSMQGAVDINKLCEEMLDNRATLQ
Nsp8:AIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAM
QRKLEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNAR
DGCVPLNIIPLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDADSKIVQLSE
ISMDNSPNLAWPLIVTALRANSAVKLQ
Nsp12:SADAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCC
RFQEKDEDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPH
ISRQRLTKYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDI
LRVYANLGERVRQALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPG
SGVPVVDSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFK
YWDQTYHPNCVNCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYH
FRELGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNV
AFQTVKPGNFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTM
CDIRQLLFVVEVVDKYFDCYDGGCINANQVIVNNLDKSAGFPFNKWGKARLYYDSMSY
EDQDALFAYTKRNVIPTITQMNLKYAISAKNRARTVAGVSICSTMTNRQFHQKLLKSIA
ATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDYPKCDRAMPNMLRIMASL
VLARKHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVF
NICQAVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVDTDFVNEFYAYLRKH
FSMMILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPH
EFCSQHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPLT
KHPNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRYWEPEFYEAMYTP
HTVLQ

7.5.5.2 Gene cloning and protein expression

Genes coding of the Nsp7, Nsp8, and Nsp12 from SARS-CoV-2 (taxid: 2697049) were codon optimized for efficient expression in E. coli, and synthetized by Twist Bioscience, San Francisco CA. Cloning of the coding sequences were conducted as reported previously using ligation independent cloning (LIC) method [480]. The sequences of Nsp12, and Nsp7 were cloned into the pMCSG53 vector, that possesses an N-terminal His-Tag followed by a Tobacco Etch Mosaic Virus (TEV) protease cleavage site. The coding sequence of Nsp8 was inserted into pRSF vector to obtain untagged protein, and to pMCSG53 for purification of Nsp8 with His-Tag. The protein expression plasmids were transformed into the E. coli BL21(DE3)-Gold strain (Stratagene, San Diego CA). For purification of Nsp7/8/12, and Nsp8/12 complexes we co-expressed Nsp12 (pMCSG53), and Nsp8 (pRSF) in single cell culture. For purification of Nsp7/8 complex we co-expressed Nsp8 (pRSF), and Nsp7 (pMCSG53). We also expressed Nsp7, Nsp8, and Nsp12 using pMCSG53 vectors containing N-terminal His-Tag. The large-scale expression of the SARS-CoV-2 proteins were done in LB Lennox medium supplemented with ampicillin 150 µg/ml. The cells were grown at 37°C (190 rpm) until OD600nm was approximately equal to 1. After that, the incubator temperature was changed to 4°C to cool down the bacterial cell suspension. When the culture reached 18°C the medium was supplemented with the following

compounds at concentrations 0.2 mM IPTG, 0.1% glucose, 40mM K2HPO4. Next, the induced bacteria culture was grown for 20 hours at 16°C with 190 rpm shaking. Subsequently cells were harvested by centrifugation at 7k RCF, and cell pellets were resuspended in lysis buffer comprising 50 mM HEPES pH 8.0, 500 mM NaCl, 5% v/v glycerol, 20 mM imidazole, and 10 mM βmercaptoethanol.

Deuterium labeled Nsp7, for contrast variation SANS experiments, was expressed using a protocol established by Marley et al. [481]. Briefly, LB cultures were grown and monitored until OD600nm reached ~0.7. At that point, the cells were pelleted by centrifugation at 4000×g for 10 minutes in pre-sterilized bottles. Each pellet was then washed in deuterated minimal medium and re-pelleted by centrifugation at 4000×g for 10 minutes. The washed pellets were then resuspended in fresh deuterated minimal medium and transferred to dry, sterile 2.8L Fernbach flasks (250 mL per flask). The cells were then cultivated at 18 °C for 1 hour, induced with 0.2 mM IPTG for ~20 hours, and harvested by centrifugation at 4000×g for 30 minutes. After decanting the supernatant, the protiated and deuterated pellets were stored at -80°C until needed for subsequent protein purification.

7.5.5.3 Protein purification

Harvested E. coli cell pellets were sonicated on ice at 120W for 5 minutes (4 sec pulses of sonication followed by 20 sec breaks). Cell debris were removed through centrifugation at 30k RCF, 4°C for 1 hour. Supernatants were mixed with 5 ml of Ni2+ Sepharose (GE Healthcare Life Sciences, Marlborough MA) previously equilibrated with lysis buffer. Solutions were transferred on chromatography Flex-Column (420400-2510) attached to Vac-Man vacuum system (Promega, Madison WI). After loading the cell lysate, the nickel beads were washed three times with 60 ml of lysis buffer. As next, proteins were eluted using 20 ml of lysis buffer enriched with 500 mM imidazole pH 8.0. Protein concertation's were measured using NanoDrop (Wilmington, DE) spectrophotometer. Subsequently after determination of sample concentration the TEV protease was added in molar ratio 1:30 (TEV : protein). These solutions were incubated for 16 hours at 4°C.

To obtain the Nsp7/8/12 complex we purified Nsp8/12 complex from co-expression system, and as next we added separately purified Nsp7, His-Tags were removed from both Nsp7, and Nsp12 using TEV protease. For purification of Nsp8/12 complex we used co-expression system with a His-Tag on Nsp12. For purification of Nsp7/8 we used a co-expression system with His-Tag on Nsp7. We also purified the Nsp7, Nsp8, and Nsp12 separately using the TEV to remove N-terminal His-Tag attached to the proteins. Subsequently, all purified proteins or obtained complexes were concentrated using 30 kDa cut-of centrifugal protein concentrators (Merck-Millipore, Burlington, MA). The complex of Nsp7/8/12 was concentrated to 20 mg/ml, the Nsp7/8 was concentrated to 50 mg/ml, and Nsp8, Nsp7, Nsp8/12, Nsp12 were concentrated to approximately 30 mg/ml. Depending on the sample volume, SEC was carried out using a Superdex 200 column in lysis buffer containing 1 mM TCEP instead of β-mercaptoethanol. After SEC gel electrophoresis was done to select the pure fractions that were frozen for further analysis. For purification of Nsp7, Nsp8, Nsp12 the selected fractions after SEC were pooled together and subsequently loaded on Ni2+ Sepharose to remove TEV, and contaminating material that binds to the nickel column. Flow-through containing pure protein were

concentrated using centrifugal concentrators, and buffer was exchanged with 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, pH 7.5. Protein purity was analyzed and compared with Precision Plus Protein™ Dual Color Standard (BioRad, Hercules, CA) using SDS-PAGE in Trisglycine buffer and molecular markers. For screening crystallization conditions, we used freshly purified proteins. Proteins were flash cooled at liquid nitrogen for subsequent SAXS and SANS data collection.

7.5.5.4   Co-purification of deuterated Nsp7 and protiated Nsp8 complex

The overexpressed deuterated Nsp7 (dNSP7) and protiated Nsp8 E. coli cell pastes were resuspended in 50 mL lysis buffer containing 50 mM HEPES, 500 mM NaCl, 20 mM imidazole, 5% glycerol, 1 mM TCEP and EDTA-free protease inhibitor tablet Complete (05056489001, Roche, USA), respectively. The resuspended cells were mixed together and lysed by sonication on ice for 10 min with 50% amplitude using a 4 sec on/10 sec off cycle. The clarified lysate was loaded onto 5 mL HisTrapTM HP column (17524802, Cytiva, Marlborough MA) after centrifugation at 18,000 rpm for 30 min. The column was washed by 10 column volumes (CV) of buffer A (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, 5% glycerol, and 1mM TCEP), 10 CV of 6% buffer B (50 mM HEPES, 500 mM NaCl, 500 mM imidazole, 5% glycerol, and 1 mM TCEP) before elution. The protein was eluted with a 10 CV, 6 to 100% linear gradient. The fractions containing dNsp7 and Nsp8 were confirmed by SDS-PAGE and pooled. The pooled fractions were dialyzed against buffer containing 50 mM HEPES, 300 mM NaCl, 5% glycerol, 1 mM TCEP with TEV protease (1:50 ratio) at 4 °C for 16 hours. The cleaved protein was loaded to a 5-mL HisTrapTM HP column to remove TEV protease and other contaminants. Flow-through was collected and concentrated using a centrifugal concentrator with a 30kDa MW cutoff and then loaded to a HiLoad 16/60 Superdex 200 column equilibrated in 50 mM HEPES, 150 mM NaCl, 1 mM TCEP for final. After pooling and concentrating the peak fractions, the purity and concentration of the purified protein were determined by SDS-PAGE and UV-Vis spectrophotometry. Using dialysis, the relative ratio of D2O/H2O in the protein solution was adjusted to 90% D2O. In 90% D2O solvent, the scattering contribution of dNSP7 was negligible because the scattering length density of dNSP7 was matched to that of the solvent.

7.5.5.5   RNA extension essay

The RNA used in the extension assay folds into a hairpin with a GAAA tetraloop and a single-stranded tail, 5'-GGCUU-3', at its 5' terminus (Figure 7-40). Its sequence is composed of 31 nucleotides: 5'pppGGCUUAGGAGAUGAUGAAAGUCAUUCUCCU-OH-3'. We synthesized the 31-mer by in vitro transcription as described previously [482]. Oligonucleotides used for the construction of the DNA template are:

5'-AATTCCTGCAGTAATACGACTCACTATAGGCTTAGGAGAATGATGAAAGTC-3' and 5'mAmGGAGAATGACTTTCATCATTCT-3'.

mA and mG represent the 2'-O-methyl nucleotides used to reduce the proportion of runover transcripts. The underlined portions represent the overlapped segment and bold type indicates the T7 RNA polymerase promoter. To synthesize the internally [32P]-labeled 31-mer transcripts, 10 ug/ml duplex DNA in a solution containing 200 mM HEPES/KOH (pH7.5), 30 mM MgCl2, 2 mM spermidine, 40 mM DTT, 3 mM each ATP, CTP, GTP and UTP, 0.33 µM [alpha-32P]ATP was incubated with 100 µg/mL T7 RNA polymerase, 5 U/ml inorganic phosphatase (New England BioLabs, Boston MA), SUPERase. In RNase Inhibitor (Invitrogen, Walton MA) in a 50 µL reaction for 5 hours at 37 °C. The DNA template was removed by addition of 50 U/ml RQ1 RNase-free DNase (Promega, Madison WI) and incubation for 15 min at 37 °C, followed by phenol-chloroform extraction. The 31-mer RNA was recovered from the aqueous phase by ethanol precipitation and then purified by electrophoresis on a 15% denaturing polyacrylamide gel. To remove homodimers and produce a homogeneous solution of hairpins, the gel-purified RNA transcripts dissolved in water were heated to 75 °C and gradually cooled to room temperature. The RNA extension reactions contained 10 µM 31-mer RNA hairpin, 5 µM Nsp12, 10 µM Nsp8, 5 µM Nsp7 in 20 mM HEPES/KOH (pH 7.5), 75 mM NaCl, 2 mM MgCl2 and 0.5 mM TCEP. Reactions were incubated for 30 min at 30 °C and the RNA extension was initiated by addition of 1 mM each ATP, CTP, GTP and UTP. Reactions were stopped by the addition of 2.5 mM EDTA (pH 8.0) followed by phenolchloroform extraction. RNA products of extension reactions were recovered from the aqueous phase by ethanol precipitation, analyzed by electrophoresis on a 15% denaturing polyacrylamide gel and visualized by autoradiography.

7.5.5.6   RNA and DNA constructs

long ssRNA: 36 bases (5'-UUU UCA UGC UAC GCG UAG CAU GCU ACG CGU AGC AUG-3')

short ssRNA: 28 bases (5'-CAU GCU ACG CGU AGC AUG CUA CGC GUA G-3')

The RNA and DNA constructs were highly purified synthetic polynucleotides ordered from IDT. The dsRNA consisted of duplexed long ssRNA (36 bases) paired to short ssRNA (28 bases). RNA oligos were dissolved and annealed in water. Both RNAs were validated by the SEC-MALS-SAXS experiment on a Shodex KW403 (Shodex, Tokyo Japan) column using 50 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM TCEP running buffer. Both MALS and SAXS measurements agree with models of monodispersed dsRNA and ssRNA (Figure 7-42). The atomistic models of both RNA's give an excellent match to the SAXS data (Figure 7-42C).

7.5.5.7   Crystallization, data collection, and structural analysis of Nsp7/8 complex

For crystallizations trials we used 400 nL protein solution mixed with 400 nL of buffer reservoir solution. The sitting drops were equilibrated against 130 µL of reservoir solution. Screening of crystallization condition of Nsp7/Nsp8 complex were performed at 289K with MCSG1, MCSG2, MCSG3, MCSG4, INDEX, Natrix HT crystallizations screens (Anatrace, Hampton Research, Maumee, OH; Hampton Research, Aliso Viejo, CA). NSP7/8A (PDB: 6WIQ) crystals were obtain from 20 mg/ml protein concentration in crystallization condition 0.1 M Tris pH 8.5, 1.5 M

ammonium phosphate dibasic. Crystals suitable for high resolution diffraction data collection appeared after three weeks. For Nsp7/8B (PDB entry 6WQD) the crystal of the Nsp7/8 was obtained using 40 mg/ml of the complex in the buffer: 0.2 M magnesium chloride, 0.1 M Tris, pH 8.5, 20% w/v PEG8000. The Nsp7/8C (PDB entry 6XIP) crystal grows from 40 mg/ml in the buffer: 0.2 M magnesium chloride hexahydrate, 0.1 M BIS-TRIS pH 5.5, 25% w/v polyethylene glycol 3350. The Nsp7/8B and C (PDB entries 6WQD and 6XIP) crystals were both obtained in a presence of V8 protease. Before setting the crystallization plates to purified Nsp7/8 we added endo-proteinase Glu-C (V8 protease) from Staphylococcus aureus (Sigma Aldrich, St. Louis MO). We used 300 times excess of Nsp7/8 over V8 protease. After 16 hours of incubation of the Nsp7/8 complex with V8 protease we set up the crystallization plates.

Prior to data collection at 100 K, all cryoprotected crystals of Nsp7-Nsp8 complex were flash-cooled in liquid nitrogen. The x-ray diffraction experiments were carried out at the Structural Biology Center 19-ID beamline at the Advanced Photon Source, Argonne National Laboratory. The diffraction images were recorded from the crystals of three different forms Nsp7/8A, Nsp7/8B, and Nsp7/8C on the PILATUS3X 6M detector using 0.3-0.5° rotation on ω and 0.5 sec exposure for 210°, 375°, and 260°, and to resolution of 2.9 Å, 1.95 Å, and 1.50 Å, respectively. For Nsp7/8B crystals, the diffractions were collected from three different spots from a crystal. The data sets were processed and scaled with the HKL3000 suite [483]. Intensities were converted to structure factor amplitudes in the Ctruncate program [484,485] from the CCP4 package [486]. The structures were determined using molrep [487] implemented in the HKL3000 software package using the SARS-CoV Nsp7/8 complex structure (PDB id 5F22) as a search model. Initial models were refined as rigid bodies and then refined all atoms by 12 cycles of REFMAC [486,488] before they were iteratively refined using COOT and PHENIX [489]. Throughout the refinement, the same 5% of reflections were kept out from the refinement in both REFMAC and PHENIX refinement. The final structures converged to Rwork = 0.218 and Rfree = 0.252 for Nsp7/8A, Rwork = 0.187 and Rfree = 0.229 for Nsp7/8B, Rwork = 0.161 and Rfree = 0.199 for Nsp7/8C with regards to each data quality. The stereochemistry of the structures were checked with PROCHECK [490] and the Ramachandran plot and validated with the PDB validation server. The data collection and processing statistics are given in Table 7-7. The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession code 6WIQ, 6WQD and 6XIP for Nsp7/8A, Nsp7/8B, and Nsp7/8C, respectively.

7.5.5.8 Size-exclusion chromatography coupled with multi-angle light scattering and small-angle X-ray scattering (SEC-MALS-SAXS)

For SEC-MALS-SAXS experiments, 60 μL containing either 10 mg/ml Nsp7; 8 mg/ml Nsp8; 8 mg/ml Nsp7/8; 6 mg/ml Nsp8/12; 6 mg/ml and Nsp7/8/12; in 50 mM HEPES pH 7.5, 150 mM NaCl, 1mM TCEP were utilized. Additionally, we also performed SEC-MALS-SAXS for the mixture of dsRNA or ssRNA with the Nsp7, Nsp8, Nsp7/8, Nsp12/8, and Nsp7/8/12 in 1:1 molar ratio.

SEC-MALS-SAXS data were collected at the Advanced Light Source (ALS) beamline SIBYLS (beamline 12.3.1) in Berkeley, California [2,9]. The X-ray wavelength was set at λ=1.127 Å, and the sample-to-detector distance was 2070 mm, resulting in scattering vectors, q, ranging from

0.01 Å^{-1} to 0.35 Å^{-1}. The scattering vector is defined as q = 4πsinθ/λ, where 2θ is the scattering angle. All experiments were performed at 20°C and data was processed as described [6]. Briefly, a SAXS flow cell (see Section 2.3.2) was coupled in-line with an Agilent 1290 Infinity HPLC system using a Shodex KW803 column equilibrated at a 0.5 mL/min flow rate with the running buffer as indicated above. 55 µL of each sample was run through the SEC column and 2s X-ray exposures were collected continuously during a 20-minute elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by the radius of gyration (Rg) derived by the Guinier approximation I(q) = I(0) exp(q2Rg2/3) with the limits q Rg<1.5 [491]. The elution peak was mapped by comparing the integral ratios to background and Rg relative to the recorded frame using the program SCÅTTER. Non-uniform Rg values across an elution peak represent a heterogeneous assembly. Final merged SAXS profiles, derived by integrating multiple frames at the peak of the elution peak or further indicated, were used for analysis, including the Guinier plot, which determined aggregation-free state. The program SCÅTTER was used to compute the pair distribution function (P(r)). The distance r where P(r) approaches zero intensity identifies the maximal dimension of the macromolecule (Dmax) (Table 7-8). P(r) functions were normalized based on the molecular weight of the assemblies as determined by SCÅTTER using the volume of correlation Vc [334] (Table 7-8). The SAXS flow-cell was additionally connected inline to a 1290 series UV-vis diode array detector (DAD) measuring at 280 and 260 nm (Agilent), 18-angle DAWN HELEOS II multi-angle light scattering (MALS) and quasi-elastic light scattering (QELS), and Optilab rEX refractometer (Wyatt Technology, Santa Barbra CA). System normalization and calibration were performed with a BSA monomer using a 45 µL sample at 10 mg/mL in the same running buffer and a dn/dc value of 0.175. The light scattering experiments were used to perform analytical scale chromatographic separations for mass and hydrodynamic radius (Rh) determination. UV, MALS, and differential refractive index data were analyzed using Wyatt Astra 7 software to monitor sample homogeneity across the elution peak complementary to the above-mentioned SEC-SAXS signal validation. The SEC-MALS signal was used as an indicator of homogeneous vs. heterogeneous samples.

7.5.5.9 High-throughput SAXS and SANS

SAXS data were collected in "batch" high throughput mode (HT-SAXS) at the ALS beamline 12.3.1 (SIBYLS) at LBNL Berkeley [333] on Nsp8 at 10, 5 and 2.5 mg/mL because of the concentration dependence we observed in SEC-SAXS and differences from SANS data. Experiments were performed 20^{0}C as described elsewhere [9]. Briefly, the sample was exposed for 10 s with the detector framing at 0.3 s to maximize signal while ensuring only non-radiation damaged signal is included in the data for analysis. Once radiation damage data was removed and a SAXS profile was integrated, processing occurred using the SCÅTTER package as described for the SEC-SAXS profiles. The protein concentrations for SANS measurements were 1.5 mg/mL Nsp7/8; 4 mg/mL Nsp8; 4 mg/mL deuterated Nsp7/Nsp8 complex, and mixture of Nsp7/8/dsRNA in 1:1 molar ratio, at 3.5 mg/mL Nsp7/8 complex in the same buffer as was performed SAXS experiments. For the contrast matching SANS experiments of deuterated dNsp7 with protiated Nsp8 (dNsp7/Nsp8 complex), 90% D2O buffer was used to selectively

highlight the scattering from Nsp8. Studies of the Nsp7/8/dsRNA mixture were performed in 65% D2O buffer to selectively highlight the scattering from Nsp7/8 complex. SANS measurements were collected at the Bio-SANS and EQ-SANS instruments located at the High Flux Isotope Reactor and Spallation Neutron Source (SNS), respectively, at Oak Ridge National Laboratory [492,493]. At the Bio-SANS, a single configuration of the dual detector system was used with the main detector at 7m (SSD) and the wing detector at 3.2°. Using this configuration, the q range spanning 0.007 < q (Å-1) < 1 was obtained using 6 Å wavelength neutrons with an $(\Delta\lambda/\lambda)$ of 15%. At EQ-SANS, two instrument configurations were used - 2.5Å and 2.5m (SSD) for high-q and 10Å and 4m (SSD) for low-q to provide an q-range of 0.006 to 1.2 Å-1. The data were corrected for instrument background, detector sensitivity, and instrument geometry using facility data reduction software, drt-SANS. All SANS measurements were performed in 1 mm path length cylindrical quartz cuvettes (Hellma, Müllheim, Germany) at 10°C. Initial SANS data analysis, including Guinier fits and pair-distribution calculations, were performed using the BioXTAS RAW program and ATSAS suite [494,495]. The pair distance distribution function (P(r)) was calculated using the indirect Fourier transform method implemented in the program GNOM[44]. Scattering data over the range 0.007 < q (Å−1) < 8/Rg were used for P(r) analysis and subsequent modeling. The SAXS-derived molecular weight was determined using the volume of Porod method as implemented in RAW [496].

7.5.5.10 Solution Structure Modeling

Nsp7: To fit experimental SEC-SAXS curves of Nsp7 dimer, we initially modeled an Nsp7 dimer by adding missing C and N-terminal regions into our Nsp7 structure taken from the Nsp7/8 crystal structure (PDBID: 6WIQ) by MODELLER[301]. To test whether the discrepancy between SAXS and Nsp7 dimer structure was due to the flexibility of exposed C-terminal helix (68-86), we perform rigid body modeling using BILBOMD[283]. In this step, the disulfide bond (Cys8-Cys8) was preserved. The experimental SAXS was then compared to theoretical scattering curves generated from atomistic models using the FOXS [291,336] and followed by multistate model selection by MultiFOXS[284]. The bad SAXS fit leads us to search for the alternative Nsp7 dimer by bypassing disulfide bond restraints. We performed SAXS based docking of two Nsp7 monomers using FOXSDock approach[479]. The best score model was superimposed on the average SAXS envelope calculated by GASBOR with a P2 symmetry operator[497] (Figure 7-36C).

Nsp8 monomer: To find an Nsp8 conformation that fits the SEC-SAXS Data, we built an Nsp8 monomer by adding missing C and N-terminal regions into the Nsp8 monomer, taken from Nsp7/8/12/RNA structure (PDBID: 6YYT)[471]. We employed a conformational sampling of the N-terminal helix-bundle region using BILBOMD[283]. The flexible tethers in between the head and two distinct helix regions (1-82 and 86-100) were identified by structural comparison of the two Nsp8 conformers from the Nsp7/8/12/RNA complex[471]. The experimental SAXS profiles were then compared to theoretical scattering curves generated from atomistic models using the FOXS [291,336] followed by multistate model selection by MultiFoXS [284]. Despite providing a nearly

exhaustive search of extended and compact conformations, an excellent fit to the SAXS data was obtained for a single closed state (Figure 7-36D).

Nsp8 dimer and tetramer: To find a multistate model that fits a higher oligomerization state collected in HT-SAXS mode or at the peak of SEC elution, we built the Nsp8 dimer and tetramer. Missing regions were added to the Nsp8 dimer taken from the crystal structure of the Nsp7/8 complex (PDBID: 3UB0) [473]. By keeping the dimerization interface, we used BILBOMD [283] to build an Nsp8 dimer with open and close C-terminal head regions. Additionally, we built the Nsp8 tetramer by adding missing regions into the Nsp8 tetramer taken from the crystal structure of Nsp7/8 (PDBID: 2AHM). The experimental SAXS profiles were then compared to theoretical SAXS curves generated from the pool of the monomers, dimers, and tetramer using the FOXS [291,336] followed by multistate model selection by MultiFOXS [284] (Figure 7-43).

Nsp12: To fit experimental SEC-SAXS curves of Nsp12 monomer, we add missing regions into Nsp12 structure, taken from Nsp7/8/12/RNA cryo-EM structure (PDBID: 6YYT), by using MODELLER [301]. The experimental SAXS profiles were then compared to theoretical scattering curves generated from atomistic models using the FOXS [291,336] (Figure 7-36E). The experimental SANS profile of Nsp8 was compared to theoretical SANS of an Nsp8 tetramer built using the SAXS model of Nsp8 dimer (PDBID: G1TQHV) using SAXS/REFMX to perform the quaternary structure modeling [498,499].

Nsp7/8: To find a multistate model that fit SEC-SAXS data for Nsp7/8, we initially build a complete model of Nsp7/8 heterotetramer by combining our crystal structure (PDBID: 6WIQ) and Nsp8 structure taken from Nsp7/8/12/RNA structure (PDBID: 6YYT)[471]. The experimental SAXS profile was then compared to theoretical SAXS curves generated from the pool of the Nsp8 monomers, Nsp8 dimers, Nsp7 dimers, Nsp7/8 heterotetramer, and Nsp7/8 heterodimer using the FOXS [291,336] followed by multistate model selection by MultiFOXS [284] (Figure 7-37E). The experimental SANS profile of Nsp7/8 complex was compared to theoretical SANS profile using a mixture of SAXS models of Nsp7/8 heterotetramer (PDBID: DJTRUW) and Nsp8 monomer (PDBID: IDDWOG). OLIGOMER was used to determine the volume fraction of component [500]. The experimental SANS profile of dNsp7/Nsp8 complex was compared to theoretical SANS profile using a mixture of SAXS models of Nsp8 dimers (PDBID: DJTRUW) and Nsp8 monomers (PDBID: IDDWOG). SAXS/REFMX was used to perform quaternary structure modeling to account for polydispersity of the system.

Nsp8/12 and Nsp7/8/12: To find a model that fits SEC-SAXS data of Nsp8/12 and Nsp7/8/12, we initially built a complete model of Nsp7/8/12 using the cryo-EM Nsp7/8/12/RNA structure (PDBID: 6YYT) [471]. The experimental SAXS curves were then compared to theoretical SAXS curves generated from the pool of the models: (Nsp7 dimer, Nsp8 monomers, Nsp8 dimers, Nsp12 monomer, Nsp8/12 with 1:1 and 2:1 ratios; and Nsp7/8/12 with 1:1:1, 1:2:1, 0:1:1, 0:2:1 ratio) using the FOXS [291,336] followed by multistate model selection by MultiFOXS [284]. Despite providing a large pool of the models to select a multistate model [284] an excellent fit for the SAXS data required only Nsp8/12 1:1 conformer to fit both data sets (Figure 7-38D).

Nsp8/dsRNA: To find a model that fits SEC-SAXS data of Nsp8/dsRNA complex, we built a complete model of Nsp8dsRNA using the cryo-EM Nsp7/8/12/RNA structure (PDBID: 6YYT) [471]. The flexible tethers in between the head (101-199) and helix bundle regions (1-82 and 86-100) were used to perform conformational sampling of the head region using BILBOMD [283]. The

experimental SAXS profiles were then compared to theoretical scattering curves generated from atomistic models using the FOXS [291,336] followed by multistate model selection by MultiFoXS [284]. Despite providing a nearly exhaustive search of extended and compact conformations, an excellent fit to the SAXS data was obtained for a single closed state (Figure 7-37C). The Nsp8/dsRNA conformers pool, together with the above characterized Nsp7 dimer, were used to select a multistate model by MultiFoXS [284] for Nsp7/8+ dsRNA SEC-SAXS data (Figure 7-37F).

Nsp7/8/12/dsRNA and Nsp7/8/12/ssRNA: To fit SEC-SAXS data of Nsp7/8/12/dsRNA complex, we built a complete model of Nsp7/8/12/dsRNA using cryo-EM structure (PDBID: 6YYT)[471]. Despite providing a large pool of the models to select a multistate model [284] an excellent fit for the SAXS data required only a complete complex of Nsp7/8/12/dsRNA with 1:1:2:1:1 ratio (Figure 7-38D). Nsp7/8/12/ssRNA fit SEC-SAXS data of Nsp7/8/12/ssRNA complex with 1:1:2:1:1 ratio model using longer ssRNA (Figure 7-38D). The experimental SANS profile of Nsp7/8/dsRNA mixture was compared to theoretical SANS profile using the SAXS model of Nsp8 monomer (PDBID: 7MRNJA) CRYSON [501].

8   Concluding Remarks and Suggested Future Directions

Individual chapters provide detailed conclusions for each portion of the work. Here, I briefly present prevailing messages:

Chapter 2: SEC-SAXS-MALS measurements have become widely preferred over traditional SAXS measurements for studying structural biology primarily due to its versatility and *in situ* purification. We describe the development and use of this robust technique.

Chapter 3: XSI is a powerful method for studying DNA damage repair and DNA-Protein interactions. We describe the development and successful application monitoring nuclease activity and provides insight into DNA-protein interactions for MRE11. This demonstrates how XSI can aid in the development of inhibitors that trap enzymes on the DNA substrate.

Chapter 4: We further demonstrates the utility of XSI in studying the DNA-protein dynamics by exploring the clinically important poly(ADP-ribose) polymerase 1 (PARP-1). We investigate the damage recognition, activation, and inhibition of PARP-1 as it pertains to single-strand break (SSB) repair using XSI and suggested a methodology for combining XSI with size exclusion or anion exchange coupled chromatography small-angle X-ray scattering.

Chapter 5 and 6: We show how conventional SAXS can be used to explore protein corona formation on ssDNA-SWCNTs, underlining the need for more advanced X-ray scattering techniques (Chapter 5). This leads to the development of XSI, demonstrating a novel and powerful use of the technique enabling the first direct measurement of the surface adsorbed morphology of ssDNA on the SWCNT surface at biologically applicable concentrations in solution. These findings provide insights into the nanosensing mechanisms of ssDNA-SWCNTs and pave the way towards the development of future design strategies for DNA-functionalized SWCNT technologies.

Chapter 7: We demonstrate four case studies for SEC-SAXS-MALS system selected to best highlight the robust capabilities of this technique bringing together concepts from previous chapters.

X-ray scattering has been and continues to be a powerful technique for exploring structural biology and aiding in the rational design of bionanotechnologies. HT-XSI and SEC-SAXS-MALS techniques as well as other HT-SAXS measurement platforms have been a primary focus of the SIBYLS beamline over the last five years and continues to be a point of active development. As we continue to refine our techniques and streamline our data analysis pipeline, we look to the future of brighter light sources, improved sample delivery options, and advanced computational tools which may be the next revolution in X-ray scattering.

9   Appendix – Additional Dissertation Metadata

Figure 9-1. Dissertation metadata as a function of number of days including number of pages written (orange), weight in pounds (green), distance rowed on the rowing machine in miles (blue), and finally, in order to stave off loneliness, number of Star Trek episodes watched in the background while writing.

Figure 9-2. Word cloud of most commonly used words in dissertation quantified as word size.

266

References

1. Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly Reviews of Biophysics *40*. 10.1017/S0033583507004635.

2. Classen, S., Hura, G.L., Holton, J.M., Rambo, R.P., Rodic, I., McGuire, P.J., Dyer, K., Hammel, M., Meigs, G., Frankel, K.A., and Tainer, J.A. (2013). Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. Journal of Applied Crystallography *46*, 1-13. 10.1107/S0021889812048698.

3. Foster, B.M., Rosenberg, D., Salvo, H., Stephens, K.L., Bintz, B.J., Hammel, M., Ellenberger, T., Gainey, M.D., and Wallen, J.R. (2019). Combined Solution and Crystal Methods Reveal the Electrostatic Tethers That Provide a Flexible Platform for Replication Activities in the Bacteriophage T7 Replisome. Biochemistry *58*, 4466-4479. 10.1021/acs.biochem.9b00525.

4. Horst, B.G., Yokom, A.L., Rosenberg, D.J., Morris, K.L., Hammel, M., Hurley, J.H., and Marletta, M.A. (2019). Allosteric activation of the nitric oxide receptor soluble guanylate cyclase mapped by cryo-electron microscopy. eLife *8*, e50634. 10.7554/eLife.50634.

5. Hura, G.L., Hodge, C.D., Rosenberg, D., Guzenko, D., Duarte, J.M., Monastyrskyy, B., Grudinin, S., Kryshtafovych, A., Tainer, J.A., Fidelis, K., and Tsutakawa, S.E. (2019). Small angle X-ray scattering-assisted protein structure prediction in CASP13 and emergence of solution structure differences. Proteins: Structure, Function, and Bioinformatics *87*, 1298-1314. 10.1002/prot.25827.

6. Knott, G.J., Cress, B.F., Liu, J.J., Thornton, B.W., Lew, R.J., Al-Shayeb, B., Rosenberg, D.J., Hammel, M., Adler, B.A., Lobba, M.J., et al. (2019). Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. Elife *8*. 10.7554/eLife.49110.

7. Curran, S.C., Pereira, J.H., Baluyot, M.-J., Lake, J., Puetz, H., Rosenburg, D.J., Adams, P., and Keasling, J.D. (2020). Structure and Function of BorB, the Type II Thioesterase from the Borrelidin Biosynthetic Gene Cluster. Biochemistry *59*, 1630-1639. 10.1021/acs.biochem.0c00126.

8. Topolska-Woś, A.M., Sugitani, N., Cordoba, J.J., Le Meur, K.V., Le Meur, R.A., Kim, H.S., Yeo, J.-E., Rosenberg, D., Hammel, M., Schärer, O.D., and Chazin, W.J. (2020). A key interaction with RPA orients XPA in NER complexes. Nucleic Acids Research *48*, 2173-2188. 10.1093/nar/gkz1231.

9. Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., 2nd, Tsutakawa, S.E., Jenney, F.E., Jr., Classen, S., Frankel, K.A., Hopkins, R.C., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods *6*, 606-612.

10. Dyer, K.H., Michal; Rambo, Robert P.; Tsutakawa, Susan E.; Rodic, Ivan; Classen, Scott; Tainer, John A.; Hura, Greg L. (2013). High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. Methods in molecular biology (Clifton, N.J.) *1091*, 245-258. 10.1007/978-1-62703-691-7_18.

11. Hura, G.L.T., Chi Lin; Claridge, Shelley A.; Mendillo, Marc L.; Smith, Jessica M.; Williams, Gareth J.; Mastroianni, Alexander J.; Alivisatos, A. Paul; Putnam, Christopher D.; Kolodner, Richard D.; Tainer, John A. (2013). DNA conformations in mismatch repair probed in solution by

X-ray scattering from gold nanocrystals. Proceedings of the National Academy of Sciences of the United States of America *110*, 17308-17313. 10.1073/pnas.1308595110.

12. Rosenberg, D.J., Syed, A., Tainer, J.A., and Hura, G.L. (2022). Monitoring Nuclease Activity by X-Ray Scattering Interferometry Using Gold Nanoparticle-Conjugated DNA. In DNA Damage Responses, N. Mosammaparast, ed. (Springer US), pp. 183-205.

13. Mathew, E., Mirza, A., and Menhart, N. (2004). Liquid-chromatography-coupled SAXS for accurate sizing of aggregating proteins. Journal of Synchrotron Radiation *11*, 314-318. 10.1107/S0909049504014086.

14. Watanabe, Y., and Inoko, Y. (2009). Size-exclusion chromatography combined with small-angle X-ray scattering optics. Journal of Chromatography A *1216*, 7461-7465. 10.1016/j.chroma.2009.02.053.

15. Berthaud, A., Manzi, J., Pérez, J., and Mangenot, S. (2012). Modeling Detergent Organization around Aquaporin-0 Using Small-Angle X-ray Scattering. Journal of the American Chemical Society *134*, 10080-10088. 10.1021/ja301667n.

16. Graewert, M.A., Franke, D., Jeffries, C.M., Blanchet, C.E., Ruskule, D., Kuhle, K., Flieger, A., Schäfer, B., Tartsch, B., Meijers, R., and Svergun, D.I. (2015). Automated Pipeline for Purification, Biophysical and X-Ray Analysis of Biomacromolecular Solutions. Scientific Reports *5*, 10734. 10.1038/srep10734.

17. Pérez, J., and Vachette, P. (2017). A Successful Combination: Coupling SE-HPLC with SAXS. In Biological Small Angle Scattering: Techniques, Strategies and Tips, B. Chaudhuri, I.G. Muñoz, S. Qian, and V.S. Urban, eds. (Springer Singapore), pp. 183-199.

18. Bucciarelli, S., Midtgaard, S.R., Nors Pedersen, M., Skou, S., Arleth, L., and Vestergaard, B. (2018). Size-exclusion chromatography small-angle X-ray scattering of water soluble proteins on a laboratory instrument. Journal of Applied Crystallography *51*, 1623-1632. 10.1107/S1600576718014462.

19. Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Current Opinion in Structural Biology *58*, 197-213. 10.1016/j.sbi.2019.04.004.

20. Inoue, R., Nakagawa, T., Morishima, K., Sato, N., Okuda, A., Urade, R., Yogo, R., Yanaka, S., Yagi-Utsumi, M., Kato, K., et al. (2019). Newly developed Laboratory-based Size exclusion chromatography Small-angle x-ray scattering System (La-SSS). Scientific Reports *9*, 12610. 10.1038/s41598-019-48911-w.

21. Graewert, M.A., Da Vela, S., Gräwert, T.W., Molodenskiy, D.S., Blanchet, C.E., Svergun, D.I., and Jeffries, C.M. (2020). Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data. Crystals *10*, 975. 10.3390/cryst10110975.

22. Chen, Z., Kibler, R.D., Hunt, A., Busch, F., Pearl, J., Jia, M., VanAernum, Z.L., Wicky, B.I.M., Dods, G., Liao, H., et al. (2020). De novo design of protein logic gates. Science *368*, 78-84. 10.1126/science.aay2790.

23. Wei, K.Y., Moschidi, D., Bick, M.J., Nerli, S., McShan, A.C., Carter, L.P., Huang, P.-S., Fletcher, D.A., Sgourakis, N.G., Boyken, S.E., and Baker, D. (2020). Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. Proceedings of the National Academy of Sciences *117*, 7208-7215. 10.1073/pnas.1914808117.

24.   Faraday, M. (1857). X. The Bakerian Lecture. —Experimental relations of gold (and other metals) to light. Philosophical Transactions of the Royal Society of London *147*, 145-181. doi:10.1098/rstl.1857.0011.

25.   Gold Clusters, Colloids and Nanoparticles I. (2014). (Springer International Publishing).

26.   Piella, J., Bastús, N.G., and Puntes, V. (2016). Size-Controlled Synthesis of Sub-10-nanometer Citrate-Stabilized Gold Nanoparticles and Related Optical Properties. Chemistry of Materials *28*, 1066-1075. 10.1021/acs.chemmater.5b04406.

27.   Alivisatos, A.P., Johnsson, K.P., Peng, X., Wilson, T.E., Loweth, C.J., Bruchez, M.P., and Schultz, P.G. (1996). Organization of 'nanocrystal molecules' using DNA. Nature *382*, 609-611. 10.1038/382609a0.

28.   Mirkin, C.A., Letsinger, R.L., and Mucic, R.C. (1996). A DNA-based method for rationally assembling nanoparticles into macroscopic materials. Nature *382*, 607-609.

29.   Mathew-Fenn, R.S., Das, R., Silverman, J.A., Walker, P.A., and Harbury, P.A.B. (2008). A Molecular Ruler for Measuring Quantitative Distance Distributions. PLoS ONE *3*, e3229. 10.1371/journal.pone.0003229.

30.   Mathew-Fenn, R.S., Das, R., and Harbury, P.A.B. (2008). Remeasuring the Double Helix. Science *322*, 446-449. 10.1126/science.1158881.

31.   Schaaff, T.G., Knight, G., Shafigullin, M.N., Borkman, R.F., and Whetten, R.L. (1998). Isolation and Selected Properties of a 10.4 kDa Gold:Glutathione Cluster Compound. The Journal of Physical Chemistry B *102*, 10643-10646. 10.1021/jp9830528.

32.   Ackerson, C.J., Jadzinsky, P.D., and Kornberg, R.D. (2005). Thiolate Ligands for Synthesis of Water-Soluble Gold Clusters. Journal of the American Chemical Society *127*, 6550-6551. 10.1021/ja046114i.

33.   Glatter, O. (1977). A new method for the evaluation of small-angle scattering data. Journal of Applied Crystallography *10*, 415-421. 10.1107/S0021889877013879.

34.   Hura, G.L., Tsai, C.L., Claridge, S.A., Mendillo, M.L., Smith, J.M., Williams, G.J., Mastroianni, A.J., Alivisatos, A.P., Putnam, C.D., Kolodner, R.D., and Tainer, J.A. (2013). DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. Proc Natl Acad Sci U S A *110*, 17308-17313. 10.1073/pnas.1308595110.

35.   Loweth, C.J., Caldwell, W.B., Peng, X., Alivisatos, A.P., and Schultz, P.G. (1999). DNA-Based Assembly of Gold Nanocrystals. Angewandte Chemie International Edition *38*, 1808-1812. 10.1002/(SICI)1521-3773(19990614)38:12<1808::AID-ANIE1808>3.0.CO;2-C.

36.   Li, Z., Jin, R., Mirkin, C.A., and Letsinger, R.L. (2002). Multiple thiol-anchor capped DNA-gold nanoparticle conjugates. Nucleic Acids Research *30*, 1558-1562. 10.1093/nar/30.7.1558.

37.   Zettl, T., Mathew, R.S., Shi, X., Doniach, S., Herschlag, D., Harbury, P.A.B., and Lipfert, J. (2018). Gold nanocrystal labels provide a sequence–to–3D structure map in SAXS reconstructions. Science Advances *4*, eaar4418. 10.1126/sciadv.aar4418.

38.   Zettl, T., Mathew, R.S., Seifert, S., Doniach, S., Harbury, P.A.B., and Lipfert, J. (2016). Absolute Intramolecular Distance Measurements with Angstrom-Resolution Using Anomalous Small-Angle X-ray Scattering. Nano Letters *16*, 5353-5357. 10.1021/acs.nanolett.6b01160.

39. Zettl, T., Das, R., Harbury, P.A.B., Herschlag, D., Lipfert, J., Mathew, R.S., and Shi, X. (2018). Recording and Analyzing Nucleic Acid Distance Distributions with X-Ray Scattering Interferometry (XSI). Current Protocols in Nucleic Acid Chemistry *73*, e54. 10.1002/cpnc.54.

40. Aubin-Tam, M.-E., and Hamad-Schifferli, K. (2005). Gold Nanoparticle−Cytochrome c Complexes: The Effect of Nanoparticle Ligand Charge on Protein Structure. Langmuir *21*, 12080-12084. 10.1021/la052102e.

41. Aubin-Tam, M.-E., Hwang, W., and Hamad-Schifferli, K. (2009). Site-directed nanoparticle labeling of cytochrome c. Proceedings of the National Academy of Sciences *106*, 4095-4100. 10.1073/pnas.0807299106.

42. Rosenberg, D.J., Hura, G.L., and Hammel, M. (2022). Size exclusion chromatography coupled small angle X-ray scattering with tandem multiangle light scattering at the SIBYLS beamline. In Methods in Enzymology, (Elsevier), pp. 191-219.

43. Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole II, F.L., Tsutakawa, S.E., Jenney Jr, F.E., Classen, S., Frankel, K.A., Hopkins, R.C., et al. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nature Methods *6*, 606-612. 10.1038/nmeth.1353.

44. Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of Applied Crystallography *25*, 495-503. 10.1107/S0021889892001663.

45. Cowieson, N.P., Edwards-Gayle, C.J.C., Inoue, K., Khunti, N.S., Doutch, J., Williams, E., Daniels, S., Preece, G., Krumpa, N.A., Sutter, J.P., et al. (2020). Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source. Journal of Synchrotron Radiation *27*, 1438-1446. 10.1107/S1600577520009960.

46. Da Vela, S., and Svergun, D.I. (2020). Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution. Current Research in Structural Biology *2*, 164-170. 10.1016/j.crstbi.2020.08.004.

47. Trewhella, J. (2022). Recent advances in small-angle scattering and its expanding impact in structural biology. Structure *30*, 15-23. 10.1016/j.str.2021.09.008.

48. Graewert, M.A., and Jeffries, C.M. (2017). Sample and Buffer Preparation for SAXS. In Biological Small Angle Scattering: Techniques, Strategies and Tips, B. Chaudhuri, I.G. Muñoz, S. Qian, and V.S. Urban, eds. (Springer Singapore), pp. 11-30.

49. Hammel, M., Rosenberg, D.J., Bierma, J., Hura, G.L., Thapar, R., Lees-Miller, S.P., and Tainer, J.A. (2020). Visualizing functional dynamicity in the DNA-dependent protein kinase holoenzyme DNA-PK complex by integrating SAXS with cryo-EM. Progress in Biophysics and Molecular Biology, S0079610720300912. 10.1016/j.pbiomolbio.2020.09.003.

50. Tsutakawa, S.E., Sarker, A.H., Ng, C., Arvai, A.S., Shin, D.S., Shih, B., Jiang, S., Thwin, A.C., Tsai, M.S., Willcox, A., et al. (2020). Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. Proc Natl Acad Sci U S A *117*, 14127-14138. 10.1073/pnas.1921311117.

51. Hodge, C.D., Rosenberg, D.J., Grob, P., Wilamowski, M., Joachimiak, A., Hura, G.L., and Hammel, M. (2021). Rigid monoclonal antibodies improve detection of SARS-CoV-2 nucleocapsid protein. mAbs *13*, 1905978. 10.1080/19420862.2021.1905978.

52. Wilamowski, M., Hammel, M., Leite, W., Zhang, Q., Kim, Y., Weiss, K.L., Jedrzejczak, R., Rosenberg, D.J., Fan, Y., Wower, J., et al. (2021). Transient and stabilized complexes of Nsp7, Nsp8, and Nsp12 in SARS-CoV-2 replication. Biophysical Journal *120*, 3152-3165. 10.1016/j.bpj.2021.06.006.

53. Hammersley, A.P. (1997). FIT2D: An Introduction and Overview'. ESRF Internal Report *ESRF97HA02T*.

54. Konarev, P.V., Graewert, M.A., Jeffries, C.M., Fukuda, M., Cheremnykh, T.A., Volkov, V.V., and Svergun, D.I. (2022). EFAMIX , a tool to decompose inline chromatography SAXS data from partially overlapping components. Protein Science *31*, 269-282. 10.1002/pro.4237.

55. Hopkins, J.B., Gillilan, R.E., and Skou, S. (2017). *BioXTAS RAW* : improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. Journal of Applied Crystallography *50*, 1545-1553. 10.1107/S1600576717011438.

56. Panjkovich, A., and Svergun, D.I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. Bioinformatics *34*, 1944-1946. 10.1093/bioinformatics/btx846.

57. Meisburger, S.P., Xu, D., and Ando, N. (2021). *REGALS* : a general method to deconvolve X-ray scattering data from evolving mixtures. IUCrJ *8*, 225-237. 10.1107/S2052252521000555.

58. Guinier, A. (1939). La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. Annales de Physique *11*, 161-237. 10.1051/anphys/193911120161.

59. Meisburger, S.P., Taylor, A.B., Khan, C.A., Zhang, S., Fitzpatrick, P.F., and Ando, N. (2016). Domain Movements upon Activation of Phenylalanine Hydroxylase Characterized by Crystallography and Chromatography-Coupled Small-Angle X-ray Scattering. Journal of the American Chemical Society *138*, 6506-6516. 10.1021/jacs.6b01563.

60. Brookes, E., Vachette, P., Rocco, M., and Pérez, J. (2016). *US-SOMO* HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. Journal of Applied Crystallography *49*, 1827-1841. 10.1107/S1600576716011201.

61. Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. General physiology and biophysics *28*, 174-189.

62. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Research *44*, W424-W429. 10.1093/nar/gkw389.

63. Schneidman-Duhovny, D., and Hammel, M. (2018). Modeling Structure and Dynamics of Protein Complexes with SAXS Profiles. Methods in molecular biology (Clifton, N.J.) *1764*, 449-473. 10.1007/978-1-4939-7759-8_29.

64. Reyes, F.E., Schwartz, C.R., Tainer, J.A., and Rambo, R.P. (2014). Methods for Using New Conceptual Tools and Parameters to Assess RNA Structure by Small-Angle X-Ray Scattering. In Methods in Enzymology, (Elsevier), pp. 235-263.

65. Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477-481. 10.1038/nature12070.

66. Kok, C.M., and Rudin, A. (1981). Relationship between the Hydrodynamic Radius and the Radius of Gyration of a Polymer in Solution. Die Makromolekulare Chemie, Rapid Communications *2*, 655-659. 10.1002/marc.1981.030021102.

67. Stracker, T.H., and Petrini, J.H.J. (2011). The MRE11 complex: starting from the ends. Nature Reviews Molecular Cell Biology *12*, 90-103. 10.1038/nrm3047.

68. Syed, A., and Tainer, J.A. (2018). The MRE11–RAD50–NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. Annual Review of Biochemistry *87*, 263-294. 10.1146/annurev-biochem-062917-012415.

69. Paull, T.T. (2018). 20 Years of Mre11 Biology: No End in Sight. Molecular Cell *71*, 419-427. 10.1016/j.molcel.2018.06.033.

70. Hopfner, K.-P., Karcher, A., Craig, L., Woo, T.T., Carney, J.P., and Tainer, J.A. (2001). Structural Biochemistry and Interaction Architecture of the DNA Double-Strand Break Repair Mre11 Nuclease and Rad50-ATPase. Cell *105*, 473-485. 10.1016/S0092-8674(01)00335-X.

71. Williams, R.S., Moncalian, G., Williams, J.S., Yamada, Y., Limbo, O., Shin, D.S., Groocock, L.M., Cahill, D., Hitomi, C., Guenther, G., et al. (2008). Mre11 Dimers Coordinate DNA End Bridging and Nuclease Processing in Double-Strand-Break Repair. Cell *135*, 97-109. 10.1016/j.cell.2008.08.017.

72. Park, Young B., Chae, J., Kim, Young C., and Cho, Y. (2011). Crystal Structure of Human Mre11: Understanding Tumorigenic Mutations. Structure *19*, 1591-1602. 10.1016/j.str.2011.09.010.

73. Shibata, A., Moiani, D., Arvai, A.S., Perry, J., Harding, S.M., Genois, M.M., Maity, R., van Rossum-Fikkert, S., Kertokalio, A., Romoli, F., et al. (2014). DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. Mol Cell *53*, 7-18. 10.1016/j.molcel.2013.11.003.

74. Chen, C., and Hildebrandt, N. (2020). Resonance energy transfer to gold nanoparticles: NSET defeats FRET. TrAC Trends in Analytical Chemistry *123*, 115748. 10.1016/j.trac.2019.115748.

75. Tang, H.Y.H., Tainer, J.A., and Hura, G.L. (2017). High Resolution Distance Distributions Determined by X-Ray and Neutron Scattering. In Biological Small Angle Scattering: Techniques, Strategies and Tips, B. Chaudhuri, I.G. Muñoz, S. Qian, and V.S. Urban, eds. (Springer), pp. 167-181.

76. Rambo, R.P., and Tainer, J.A. (2013). Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. Annual Review of Biophysics *42*, 415-441. 10.1146/annurev-biophys-083012-130301.

77. Rambo, R.P., and Tainer, J.A. (2011). Characterizing Flexible and Instrinsically Unstructured Biological Macromolecules by SAS using the Porod-Debye Law. Biopolymers *95*, 559-571. 10.1002/bip.21638.

78. Young, T. (1804). I. The Bakerian Lecture. Experiments and calculations relative to physical optics. Philosophical Transactions of the Royal Society of London *94*, 1-16. 10.1098/rstl.1804.0001.

79. Vainshtein, B.K., Feigin, L.A., Lvov, Y.M., Gvozdev, R.I., Marakushev, S.A., and Likhtenshtein, G.I. (1980). Determination of the distance between heavy-atom markers in haemoglobin and histidine decarboxylase in solution by small-angle X-ray scattering. FEBS Letters *116*, 107-110. 10.1016/0014-5793(80)80539-4.

80. Mastroianni, A.J., Sivak, D.A., Geissler, P.L., and Alivisatos, A.P. (2009). Probing the Conformational Distributions of Subpersistence Length DNA. Biophysical Journal *97*, 1408-1417. 10.1016/j.bpj.2009.06.031.

81. Claridge, S.A., Mastroianni, A.J., Au, Y.B., Liang, H.W., Micheel, C.M., Fréchet, J.M.J., and Alivisatos, A.P. (2008). Enzymatic Ligation Creates Discrete Multinanoparticle Building Blocks for Self-Assembly. Journal of the American Chemical Society *130*, 9598-9605. 10.1021/ja8026746.

82. Shi, X., Herschlag, D., and Harbury, P.A.B. (2013). Structural ensemble and microscopic elasticity of freely diffusing DNA by direct measurement of fluctuations. Proceedings of the National Academy of Sciences *110*, E1444-E1451. 10.1073/pnas.1218830110.

83. Shi, X., Beauchamp, K.A., Harbury, P.B., and Herschlag, D. (2014). From a structural average to the conformational ensemble of a DNA bulge. Proceedings of the National Academy of Sciences *111*, E1473-E1480. 10.1073/pnas.1317032111.

84. Shi, X., Huang, L., Lilley, D.M.J., Harbury, P.B., and Herschlag, D. (2016). The Solution Structural Ensembles of RNA Kink-turn Motifs and Their Protein Complexes. Nature chemical biology *12*, 146-152. 10.1038/nchembio.1997.

85. Tubbs, J.L., Latypov, V., Kanugula, S., Butt, A., Melikishvili, M., Kraehenbuehl, R., Fleck, O., Marriott, A., Watson, A.J., Verbeek, B., et al. (2009). Flipping of alkylated DNA damage bridges base and nucleotide excision repair. Nature *459*, 808-813. 10.1038/nature08076.

86. Eckelmann, B.J., Bacolla, A., Wang, H., Ye, Z., Guerrero, E.N., Jiang, W., El-Zein, R., Hegde, M.L., Tomkinson, A.E., Tainer, John A., and Mitra, S. (2020). XRCC1 promotes replication restart, nascent fork degradation and mutagenic DNA repair in BRCA2-deficient cells. NAR Cancer *2*, zcaa013. 10.1093/narcan/zcaa013.

87. Thapar, R. (2018). Regulation of DNA Double-Strand Break Repair by Non-Coding RNAs. Molecules *23*, 2789. 10.3390/molecules23112789.

88. Houl, J.H., Ye, Z., Brosey, C.A., Balapiti-Modarage, L.P.F., Namjoshi, S., Bacolla, A., Laverty, D., Walker, B.L., Pourfarjam, Y., Warden, L.S., et al. (2019). Selective small molecule PARG inhibitor causes replication fork stalling and cancer cell death. Nature communications *10*, 5654. 10.1038/s41467-019-13508-4.

89. Zandarashvili, L., Langelier, M.-F., Velagapudi, U.K., Hancock, M.A., Steffen, J.D., Billur, R., Hannan, Z.M., Wicks, A.J., Krastev, D.B., Pettitt, S.J., et al. (2020). Structural basis for allosteric PARP-1 retention on DNA breaks. Science *368*, eaax6367. 10.1126/science.aax6367.

90. Bacolla, A., Tainer, J.A., Vasquez, K.M., and Cooper, D.N. (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Research *44*, 5673-5688. 10.1093/nar/gkw261.

91. Bacolla, A., Ye, Z., Ahmed, Z., and Tainer, J.A. (2019). Cancer mutational burden is shaped by G4 DNA, replication stress and mitochondrial dysfunction. Progress in Biophysics and Molecular Biology *147*, 47-61. 10.1016/j.pbiomolbio.2019.03.004.

92. Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., et al. (2021). *ATSAS 3.0* : expanded functionality and new tools for small-angle scattering data analysis. Journal of Applied Crystallography *54*, 343-355. 10.1107/S1600576720013412.

93. Hottiger, M.O., Hassa, P.O., Lüscher, B., Schüler, H., and Koch-Nolte, F. (2010). Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends in Biochemical Sciences *35*, 208-219. 10.1016/j.tibs.2009.12.003.

94. Satoh, M.S., and Lindahl, T. (1992). Role of poly(ADP-ribose) formation in DNA repair. Nature *356*, 356-358. 10.1038/356356a0.

95. Eustermann, S., Videler, H., Yang, J.-C., Cole, P.T., Gruszka, D., Veprintsev, D., and Neuhaus, D. (2011). The DNA-Binding Domain of Human PARP-1 Interacts with DNA Single-Strand Breaks as a Monomer through Its Second Zinc Finger. Journal of Molecular Biology *407*, 149-170. 10.1016/j.jmb.2011.01.034.

96. Ali, A.A.E., Timinszky, G., Arribas-Bosacoma, R., Kozlowski, M., Hassa, P.O., Hassler, M., Ladurner, A.G., Pearl, L.H., and Oliver, A.W. (2012). The zinc-finger domains of PARP1 cooperate to recognize DNA strand breaks. Nature Structural & Molecular Biology *19*, 685-692. 10.1038/nsmb.2335.

97. Caldecott, K.W. (2014). Protein ADP-ribosylation and the cellular response to DNA strand breaks. DNA Repair *19*, 108-113. 10.1016/j.dnarep.2014.03.021.

98. Eustermann, S., Wu, W.-F., Langelier, M.-F., Yang, J.-C., Easton, Laura E., Riccio, Amanda A., Pascal, John M., and Neuhaus, D. (2015). Structural Basis of Detection and Signaling of DNA Single-Strand Breaks by Human PARP-1. Molecular Cell *60*, 742-754. 10.1016/j.molcel.2015.10.032.

99. Sukhanova, M.V., Abrakhi, S., Joshi, V., Pastre, D., Kutuzov, M.M., Anarbaev, R.O., Curmi, P.A., Hamon, L., and Lavrik, O.I. (2016). Single molecule detection of PARP1 and PARP2 interaction with DNA strand breaks and their poly(ADP-ribosyl)ation using high-resolution AFM imaging. Nucleic Acids Research *44*, e60-e60. 10.1093/nar/gkv1476.

100. Langelier, M.-F., Planck, J.L., Roy, S., and Pascal, J.M. (2012). Structural Basis for DNA Damage-Dependent Poly(ADP-ribosyl)ation by Human PARP-1. Science *336*, 728-732. 10.1126/science.1216338.

101. Langelier, M.-F., and Pascal, J.M. (2013). PARP-1 mechanism for coupling DNA damage detection to poly(ADP-ribose) synthesis. Current Opinion in Structural Biology *23*, 134-143. 10.1016/j.sbi.2013.01.003.

102. Neelsen, K.J., and Lopes, M. (2015). Replication fork reversal in eukaryotes: from dead end to dynamic response. Nature Reviews Molecular Cell Biology *16*, 207-220. 10.1038/nrm3935.

103. Kraus, W.L., and Lis, J.T. (2003). PARP Goes Transcription. Cell *113*, 677-683. 10.1016/S0092-8674(03)00433-1.

104. Gottschalk, A.J., Timinszky, G., Kong, S.E., Jin, J., Cai, Y., Swanson, S.K., Washburn, M.P., Florens, L., Ladurner, A.G., Conaway, J.W., and Conaway, R.C. (2009). Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. Proceedings of the National Academy of Sciences *106*, 13770-13774. 10.1073/pnas.0906920106.

105. Lord, C.J., Tutt, A.N.J., and Ashworth, A. (2015). Synthetic Lethality and Cancer Therapy: Lessons Learned from the Development of PARP Inhibitors. Annual Review of Medicine *66*, 455-470. 10.1146/annurev-med-050913-022545.

106. Ashworth, A., and Lord, C.J. (2018). Synthetic lethal therapies for cancer: what's next after PARP inhibitors? Nature Reviews Clinical Oncology *15*, 564-576. 10.1038/s41571-018-0055-6.

107. Ferrara, R., Simionato, F., Ciccarese, C., Grego, E., Cingarlini, S., Iacovelli, R., Bria, E., Tortora, G., and Melisi, D. (2018). The development of PARP as a successful target for cancer therapy. Expert Review of Anticancer Therapy *18*, 161-175. 10.1080/14737140.2018.1419870.

108. Fang, Evandro F., Scheibye-Knudsen, M., Brace, Lear E., Kassahun, H., SenGupta, T., Nilsen, H., Mitchell, James R., Croteau, Deborah L., and Bohr, Vilhelm A. (2014). Defective Mitophagy in XPA via PARP-1 Hyperactivation and NAD+/SIRT1 Reduction. Cell *157*, 882-896. 10.1016/j.cell.2014.03.026.

109. Scheibye-Knudsen, M., Mitchell, Sarah J., Fang, Evandro F., Iyama, T., Ward, T., Wang, J., Dunn, Christopher A., Singh, N., Veith, S., Hasan-Olive, Md M., et al. (2014). A High-Fat Diet and NAD + Activate Sirt1 to Rescue Premature Aging in Cockayne Syndrome. Cell Metabolism *20*, 840-855. 10.1016/j.cmet.2014.10.005.

110. Berger, N.A., Besson, V.C., Boulares, A.H., Bürkle, A., Chiarugi, A., Clark, R.S., Curtin, N.J., Cuzzocrea, S., Dawson, T.M., Dawson, V.L., et al. (2018). Opportunities for the repurposing of PARP inhibitors for the therapy of non-oncological diseases: PARP inhibitors for repurposing. British Journal of Pharmacology *175*, 192-222. 10.1111/bph.13748.

111. Kam, T.-I., Mao, X., Park, H., Chou, S.-C., Karuppagounder, S.S., Umanah, G.E., Yun, S.P., Brahmachari, S., Panicker, N., Chen, R., et al. (2018). Poly(ADP-ribose) drives pathologic α-synuclein neurodegeneration in Parkinson's disease. Science *362*, eaat8407. 10.1126/science.aat8407.

112. Lilyestrom, W., van der Woerd, M.J., Clark, N., and Luger, K. (2010). Structural and Biophysical Studies of Human PARP-1 in Complex with Damaged DNA. Journal of Molecular Biology *395*, 983-994. 10.1016/j.jmb.2009.11.062.

113. Langelier, M.-F., Planck, J.L., Roy, S., and Pascal, J.M. (2011). 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. Journal of Biological Chemistry *286*, 10690-10701. 10.1074/jbc.M110.202507.

114. Langelier, M.-F., Servent, K.M., Rogers, E.E., and Pascal, J.M. (2008). A Third Zinc-binding Domain of Human Poly(ADP-ribose) Polymerase-1 Coordinates DNA-dependent Enzyme Activation. Journal of Biological Chemistry *283*, 4105-4114. 10.1074/jbc.M708558200.

115. Tao, Z., Gao, P., Hoffman, D.W., and Liu, H.-w. (2008). Domain C of Human Poly(ADP-ribose) Polymerase-1 Is Important for Enzyme Activity and Contains a Novel Zinc-Ribbon Motif [†,‡]. Biochemistry *47*, 5804-5813. 10.1021/bi800018a.

116. Chapman, J.D., Gagné, J.-P., Poirier, G.G., and Goodlett, D.R. (2013). Mapping PARP-1 Auto-ADP-ribosylation Sites by Liquid Chromatography–Tandem Mass Spectrometry. Journal of Proteome Research *12*, 1868-1880. 10.1021/pr301219h.

117. Loeffler, P.A., Cuneo, M.J., Mueller, G.A., DeRose, E.F., Gabel, S.A., and London, R.E. (2011). Structural studies of the PARP-1 BRCT domain. BMC Structural Biology *11*, 37. 10.1186/1472-6807-11-37.

118. Mortusewicz, O., Amé, J.-C., Schreiber, V., and Leonhardt, H. (2007). Feedback-regulated poly(ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. Nucleic Acids Research *35*, 7665-7675. 10.1093/nar/gkm933.

119. Altmeyer, M., Messner, S., Hassa, P.O., Fey, M., and Hottiger, M.O. (2009). Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. Nucleic Acids Research *37*, 3723-3738. 10.1093/nar/gkp229.

120. Rudolph, J., Mahadevan, J., Dyer, P., and Luger, K. (2018). Poly(ADP-ribose) polymerase 1 searches DNA via a 'monkey bar' mechanism. eLife *7*, e37818. 10.7554/eLife.37818.

121. Rudolph, J., Mahadevan, J., and Luger, K. (2020). Probing the Conformational Changes Associated with DNA Binding to PARP1. Biochemistry *59*, 2003-2011. 10.1021/acs.biochem.0c00256.

122. Ruf, A., MURCIAt, J.M.D., MURCIAt, G.M.D., and Schulz, G.E. (1996). Structure of the catalytic fragment of poly(ADP-ribose) polymerase from chicken. Proc. Natl. Acad. Sci. USA, 5.

123. Dawicki-McKenna, Jennine M., Langelier, M.-F., DeNizio, Jamie E., Riccio, Amanda A., Cao, Connie D., Karch, Kelly R., McCauley, M., Steffen, Jamin D., Black, Ben E., and Pascal, John M. (2015). PARP-1 Activation Requires Local Unfolding of an Autoinhibitory Domain. Molecular Cell *60*, 755-768. 10.1016/j.molcel.2015.10.013.

124. Langelier, M.-F., Zandarashvili, L., Aguiar, P.M., Black, B.E., and Pascal, J.M. (2018). NAD+ analog reveals PARP-1 substrate-blocking mechanism and allosteric communication from catalytic center to DNA-binding domains. Nature Communications *9*, 844. 10.1038/s41467-018-03234-8.

125. Fiser, A., Do, R.K., and Sali, A. (2000). Modeling of loops in protein structures. Protein Science : A Publication of the Protein Society *9*, 1753-1773.

126. Tao, Z., Gao, P., and Liu, H.-w. (2009). Identification of the ADP-Ribosylation Sites in the PARP-1 Automodification Domain: Analysis and Implications. Journal of the American Chemical Society *131*, 14258-14260. 10.1021/ja906135d.

127. Barkauskaite, E., Jankevicius, G., and Ahel, I. (2015). Structures and Mechanisms of Enzymes Employed in the Synthesis and Degradation of PARP-Dependent Protein ADP-Ribosylation. Molecular Cell *58*, 935-946. 10.1016/j.molcel.2015.05.007.

128. Murai, J., Huang, S.-y.N., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. Cancer Research *72*, 5588-5599. 10.1158/0008-5472.CAN-12-2753.

129. Marchand, J.-R., Carotti, A., Passeri, D., Filipponi, P., Liscio, P., Camaioni, E., Pellicciari, R., Gioiello, A., and Macchiarulo, A. (2014). Investigating the allosteric reverse signalling of PARP inhibitors with microsecond molecular dynamic simulations and fluorescence anisotropy. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics *1844*, 1765-1772. 10.1016/j.bbapap.2014.07.012.

130. Kumar, C., P. T. V., L., and Arunachalam, A. (2020). A mechanistic approach to understand the allosteric reverse signaling by selective and trapping poly(ADP-ribose) polymerase 1 (PARP-1) inhibitors. Journal of Biomolecular Structure and Dynamics *38*, 2482-2492. 10.1080/07391102.2019.1633406.

131. Hopkins, T.A., Shi, Y., Rodriguez, L.E., Solomon, L.R., Donawho, C.K., DiGiammarino, E.L., Panchal, S.C., Wilsbacher, J.L., Gao, W., Olson, A.M., et al. (2015). Mechanistic Dissection of PARP1 Trapping and the Impact on In Vivo Tolerability and Efficacy of PARP Inhibitors. Molecular Cancer Research *13*, 1465-1477. 10.1158/1541-7786.MCR-15-0191-T.

132. Hopkins, T.A., Ainsworth, W.B., Ellis, P.A., Donawho, C.K., DiGiammarino, E.L., Panchal, S.C., Abraham, V.C., Algire, M.A., Shi, Y., Olson, A.M., et al. (2019). PARP1 Trapping by PARP Inhibitors Drives Cytotoxicity in Both Cancer Cells and Healthy Bone Marrow. Molecular Cancer Research 17, 409-419. 10.1158/1541-7786.MCR-18-0138.

133. Hura, G.L., Budworth, H., Dyer, K.N., Rambo, R.P., Hammel, M., McMurray, C.T., and Tainer, J.A. (2013). Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. Nature Methods 10, 453-454. 10.1038/nmeth.2453.

134. Jana, N.R., Gearheart, L., and Murphy, C.J. (2001). Wet Chemical Synthesis of High Aspect Ratio Cylindrical Gold Nanorods. The Journal of Physical Chemistry B 105, 4065-4067. 10.1021/jp0107964.

135. Jin, P., Dai, Z., and Chang, T. The Role of Bis(p-Sulfonatophenyl)Phenylphosphine in Stabilizing Gold Nanoparticles. 3.

136. Hura, G.L., Tsai, C.-L., Claridge, S.A., Mendillo, M.L., Smith, J.M., Williams, G.J., Mastroianni, A.J., Alivisatos, A.P., Putnam, C.D., Kolodner, R.D., and Tainer, J.A. (2013). DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. Proceedings of the National Academy of Sciences 110, 17308-17313. 10.1073/pnas.1308595110.

137. Finnigan, J.A., and Jacobs, D.J. (1971). Light scattering by ellipsoidal particles in solution. Journal of Physics D: Applied Physics 4, 72-77. 10.1088/0022-3727/4/1/310.

138. Raval, N., Maheshwari, R., Kalyane, D., Youngren-Ortiz, S.R., Chougule, M.B., and Tekade, R.K. (2019). Importance of Physicochemical Characterization of Nanoparticles in Pharmaceutical Product Development. In Basic Fundamentals of Drug Delivery, (Elsevier), pp. 369-400.

139. Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proceedings of the National Academy of Sciences 97, 8868-8873. 10.1073/pnas.160259697.

140. Steffen, J.D., McCauley, M.M., and Pascal, J.M. (2016). Fluorescent sensors of PARP-1 structural dynamics and allosteric regulation in response to DNA damage. Nucleic Acids Research, gkw710. 10.1093/nar/gkw710.

141. Cohen, B.E., Stoddard, B.L., and Koshland, D.E. (1997). Caged NADP and NAD. Synthesis and Characterization of Functionally Distinct Caged Compounds †. Biochemistry 36, 9035-9044. 10.1021/bi970263e.

142. Yu, P., Wang, Q., Zhang, L.-H., Lee, H.-C., Zhang, L., and Yue, J. (2012). A Cell Permeable NPE Caged ADP-Ribose for Studying TRPM2. PLoS ONE 7, e51028. 10.1371/journal.pone.0051028.

143. Aoyagi-Scharber, M., Gardberg, A.S., Yip, B.K., Wang, B., Shen, Y., and Fitzpatrick, P.A. (2014). Structural basis for the inhibition of poly(ADP-ribose) polymerases 1 and 2 by BMN 673, a potent inhibitor derived from dihydropyridophthalazinone. Acta Crystallographica Section F Structural Biology Communications 70, 1143-1149. 10.1107/S2053230X14015088.

144. Thorsell, A.-G., Ekblad, T., Karlberg, T., Löw, M., Pinto, A.F., Trésaugues, L., Moche, M., Cohen, M.S., and Schüler, H. (2017). Structural Basis for Potency and Promiscuity in Poly(ADP-ribose) Polymerase (PARP) and Tankyrase Inhibitors. Journal of Medicinal Chemistry 60, 1262-1271. 10.1021/acs.jmedchem.6b00990.

145. Liu, X., Atwater, M., Wang, J., and Huo, Q. (2007). Extinction coefficient of gold nanoparticles with different sizes and different capping ligands. Colloids and Surfaces B: Biointerfaces *58*, 3-7. 10.1016/j.colsurfb.2006.08.005.

146. Manson, J., Kumar, D., Meenan, B.J., and Dixon, D. (2011). Polyethylene glycol functionalized gold nanoparticles: the influence of capping density on stability in various media. Gold Bulletin *44*, 99-105. 10.1007/s13404-011-0015-8.

147. Oesterhelt, F., Rief, M., and Gaub, H.E. (1999). Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water. New Journal of Physics *1*, 6-6. 10.1088/1367-2630/1/1/006.

148. de Gennes, P.G. (1980). Conformations of Polymers Attached to an Interface. Macromolecules *13*, 1069-1075. 10.1021/ma60077a009.

149. Cruje, C., and Chithrani, D.B. (2014). Polyethylene Glycol Functionalized Nanoparticles for Improved Cancer Treatment. Reviews in Nanoscience and Nanotechnology *3*, 20-30. 10.1166/rnn.2014.1042.

150. Li, M., Jiang, S., Simon, J., Paßlick, D., Frey, M.-L., Wagner, M., Mailänder, V., Crespy, D., and Landfester, K. (2021). Brush Conformation of Polyethylene Glycol Determines the Stealth Effect of Nanocarriers in the Low Protein Adsorption Regime. Nano Letters *21*, 1591-1598. 10.1021/acs.nanolett.0c03756.

151. Schulz, L.G. (1954). The Optical Constants of Silver, Gold, Copper, and Aluminum. I. The Absorption Coefficient. *44*.

152. Schulz, L.G., and Tangherlini, F.R. (1954). Optical Constants of Silver, Gold, Copper, and Aluminum II The Index of Refraction n. Journal of the Optical Society of America *44*, 362. 10.1364/JOSA.44.000362.

153. Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., and Higdon, D. (2009). Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. International Journal of Nonlinear Sciences and Numerical Simulation *10*. 10.1515/IJNSNS.2009.10.3.273.

154. Wolfram, S. (2003). The Mathematica book: the definitive best-selling presentation of Mathematica by the creator of the system, 5. ed Edition (Wolfram Media).

155. Boyken, S.E., Benhaim, M.A., Busch, F., Jia, M., Bick, M.J., Choi, H., Klima, J.C., Chen, Z., Walkey, C., Mileant, A., et al. (2019). De novo design of tunable, pH-driven conformational changes. Science *364*, 658-664. 10.1126/science.aav7897.

156. Knott, G.J., Cress, B.F., Liu, J.-J., Thornton, B.W., Lew, R.J., Al-Shayeb, B., Rosenberg, D.J., Hammel, M., Adler, B.A., Lobba, M.J., et al. (2019). Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. eLife *8*, e49110. 10.7554/eLife.49110.

157. Liu, Z., Tabakman, S.M., Welsher, K., and Dai, H. (2009). Carbon Nanotubes in Biology and Medicine: In vitro and in vivo Detection, Imaging and Drug Delivery. Nano research *2*, 85-120. 10.1007/s12274-009-9009-8.

158. Al-Jamal, W.T., and Kostarelos, K. (2011). Liposomes: From a Clinically Established Drug Delivery System to a Nanoparticle Platform for Theranostic Nanomedicine. Accounts of chemical research *44*, 1094-1104. 10.1021/ar200105p.

159. Nel, A.E., Mädler, L., Velegol, D., Xia, T., Hoek, E.M.V., Somasundaran, P., Klaessig, F., Castranova, V., and Thompson, M. (2009). Understanding biophysicochemical interactions at the nano–bio interface. Nature materials *8*, 543-557. 10.1038/nmat2442.

160. Monopoli, M.P., Åberg, C., Salvati, A., and Dawson, K.A. (2012). Biomolecular coronas provide the biological identity of nanosized materials. Nature nanotechnology *7*, 779-786. 10.1038/nnano.2012.207.

161. Cedervall, T., Lynch, I., Foy, M., Berggård, T., Donnelly, S.C., Cagney, G., Linse, S., and Dawson, K.A. (2007). Detailed Identification of Plasma Proteins Adsorbed on Copolymer Nanoparticles. Angewandte Chemie *119*, 5856-5858. 10.1002/ange.200700465.

162. Milani, S., Bombelli, F.B., Pitek, A.S., Dawson, K.A., and Rädler, J.O. (2012). Reversible versus Irreversible Binding of Transferrin to Polystyrene Nanoparticles: Soft and Hard Corona. ACS nano *6*, 2532-2541. 10.1021/nn204951s.

163. Horbett, T.A. (2018). Fibrinogen adsorption to biomaterials. Journal of biomedical materials research. Part A *106*, 2777-2788. 10.1002/jbm.a.36460.

164. Zhang, J., Landry, M.P., Barone, P.W., Kim, J.-H., Lin, S., Ulissi, Z.W., Lin, D., Mu, B., Boghossian, A.A., Hilmer, A.J., et al. (2013). Molecular recognition using corona phase complexes made of synthetic polymers adsorbed on carbon nanotubes. Nature nanotechnology *8*, 959-968. 10.1038/nnano.2013.236.

165. Costa, P.M., Bourgognon, M., Wang, J.T.-W., and Al-Jamal, K.T. (2016). Functionalised carbon nanotubes: From intracellular uptake and cell-related toxicity to systemic brain delivery. Journal of controlled release : official journal of the Controlled Release Society *241*, 200-219. 10.1016/j.jconrel.2016.09.033.

166. Beyene, A.G. (2019). Imaging striatal dopamine release using a nongenetically encoded near infrared fluorescent catecholamine nanosensor. SCIENCE ADVANCES.

167. Kruss, S., Landry, M.P., Vander Ende, E., Lima, B.M.A., Reuel, N.F., Zhang, J., Nelson, J., Mu, B., Hilmer, A., and Strano, M. (2014). Neurotransmitter Detection Using Corona Phase Molecular Recognition on Fluorescent Single-Walled Carbon Nanotube Sensors. Journal of the American Chemical Society *136*, 713-724. 10.1021/ja410433b.

168. Beyene, A.G., Alizadehmojarad, A.A., Dorlhiac, G., Goh, N., Streets, A.M., Král, P., Vuković, L., and Landry, M.P. (2018). Ultralarge Modulation of Fluorescence by Neuromodulators in Carbon Nanotubes Functionalized with Self-Assembled Oligonucleotide Rings. Nano Letters *18*, 6995-7003. 10.1021/acs.nanolett.8b02937.

169. Galassi, T.V., Jena, P.V., Shah, J., Ao, G., Molitor, E.A., Bram, Y., Frankel, A., Park, J., Jessurun, J., Ory, D.S., et al. (2018). An optical nanoreporter of endolysosomal lipid accumulation reveals enduring effects of diet on hepatic macrophages in vivo. Science translational medicine *10*, NA-NA. 10.1126/scitranslmed.aar2680.

170. Iverson, N.M., Barone, P.W., Shandell, M., Trudel, L.J., Sen, S., Sen, F., Ivanov, V., Atolia, E., Farias, E., McNicholas, T.P., et al. (2013). In vivo biosensing via tissue-localizable near-infrared-fluorescent single-walled carbon nanotubes. Nature Nanotechnology *8*, 873-880. 10.1038/nnano.2013.222.

171. Lundqvist, M., Stigler, J., Elia, G., Lynch, I., Cedervall, T., and Dawson, K.A. (2008). Nanoparticle size and surface properties determine the protein corona with possible implications

for biological impacts. Proceedings of the National Academy of Sciences of the United States of America *105*, 14265-14270. 10.1073/pnas.0805135105.

172. Monopoli, M.P., Walczyk, D., Campbell, A., Elia, G., Lynch, I., Bombelli, F.B., and Dawson, K.A. (2011). Physical−Chemical Aspects of Protein Corona: Relevance to in Vitro and in Vivo Biological Impacts of Nanoparticles. Journal of the American Chemical Society *133*, 2525-2534. 10.1021/ja107583h.

173. Tenzer, S., Docter, D., Kuharev, J., Musyanovych, A., Fetz, V., Hecht, R., Schlenk, F., Fischer, D., Kiouptsi, K., Reinhardt, C., et al. (2013). Rapid formation of plasma protein corona critically affects nanoparticle pathophysiology. Nature nanotechnology *8*, 772-781. 10.1038/nnano.2013.181.

174. Schöttler, S., Becker, G., Winzen, S., Steinbach, T., Mohr, K., Landfester, K., Mailänder, V., and Wurm, F.R. (2016). Protein adsorption is required for stealth effect of poly(ethylene glycol)- and poly(phosphoester)-coated nanocarriers. Nature nanotechnology *11*, 372-377. 10.1038/nnano.2015.330.

175. Godin, A.G., Varela, J.A., Gao, Z., Danné, N., Dupuis, J.P., Lounis, B., Groc, L., and Cognet, L. (2017). Single-nanotube tracking reveals the nanoscale organization of the extracellular space in the live brain. Nature Nanotechnology *12*, 238-243. 10.1038/nnano.2016.248.

176. Hong, G., Diao, S., Chang, J., Antaris, A.L., Chen, C., Zhang, B., Zhao, S., Atochin, D.N., Huang, P.L., Andreasson, K.I., et al. (2014). Through-skull fluorescence imaging of the brain in a new near-infrared window. Nature Photonics *8*, 723-730. 10.1038/nphoton.2014.166.

177. Dinarvand, M., Neubert, E., Meyer, D., Selvaggio, G., Mann, F.A., Erpenbeck, L., and Kruss, S. (2019). Near-Infrared Imaging of Serotonin Release from Cells with Fluorescent Nanosensors. Nano Letters *19*, 6604-6611. 10.1021/acs.nanolett.9b02865.

178. Pinals, R.L., Yang, D., Lui, A., Cao, W., and Landry, M.P. (2020). Corona Exchange Dynamics on Carbon Nanotubes by Multiplexed Fluorescence Monitoring. Journal of the American Chemical Society *142*, 1254-1264. 10.1021/jacs.9b09617.

179. Cedervall, T., Lynch, I., Lindman, S., Berggård, T., Thulin, E., Nilsson, H., Dawson, K.A., and Linse, S. (2007). Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. Proceedings of the National Academy of Sciences of the United States of America *104*, 2050-2055. 10.1073/pnas.0608582104.

180. Dutta, A.K., Rösgen, J., and Rajarathnam, K. (2015). Using Isothermal Titration Calorimetry to Determine Thermodynamic Parameters of Protein–Glycosaminoglycan Interactions. In Glycosaminoglycans: Chemistry and Biology, K. Balagurunathan, H. Nakato, and U.R. Desai, eds. (Springer New York), pp. 315-324. 10.1007/978-1-4939-1714-3_25.

181. Gal, N., Schroffenegger, M., and Reimhult, E. (2018). Stealth Nanoparticles Grafted with Dense Polymer Brushes Display Adsorption of Serum Protein Investigated by Isothermal Titration Calorimetry. The journal of physical chemistry. B *122*, 5820-5834. 10.1021/acs.jpcb.8b02338.

182. Wertz and, C.F., and Santore, M.M. (2001). Effect of Surface Hydrophobicity on Adsorption and Relaxation Kinetics of Albumin and Fibrinogen: Single-Species and Competitive Behavior. Langmuir *17*, 3006-3016. 10.1021/la0017781.

183. Latour, R.A. (2014). The langmuir isotherm: A commonly applied but misleading approach for the analysis of protein adsorption behavior. Journal of biomedical materials research. Part A *103*, 949-958. 10.1002/jbm.a.35235.

184. Prozeller, D., Morsbach, S., and Landfester, K. (2019). Isothermal titration calorimetry as a complementary method for investigating nanoparticle–protein interactions. Nanoscale *11*, 19265-19273. 10.1039/c9nr05790k.

185. Chen, F., Wang, G., Griffin, J.I., Brenneman, B., Banda, N.K., Holers, V.M., Backos, D.S., Wu, L.-P., Moghimi, S.M., and Simberg, D. (2016). Complement proteins bind to nanoparticle protein corona and undergo dynamic exchange in vivo. Nature nanotechnology *12*, 387-393. 10.1038/nnano.2016.269.

186. Mildner, D.F.R., and Hall, P.L. (1986). Small-angle scattering from porous solids with fractal geometry. Journal of Physics D: Applied Physics *19*, 1535-1545. 10.1088/0022-3727/19/8/021.

187. Schaefer, D.W. (1989). Polymers, Fractals, and Ceramic Materials. Science *243*, 1023-1027. 10.1126/science.243.4894.1023.

188. Schmidt, P.W. (1991). Small-Angle Scattering Studies of Disordered, Porous and Fraetal Systems. J. Appl. Cryst. *24*, 414-435.

189. Ge, C., Du, J., Zhao, L., Wang, L., Liu, Y., Li, D., Yang, Y., Zhou, R., Zhao, Y., Chai, Z., and Chen, C. (2011). Binding of blood proteins to carbon nanotubes reduces cytotoxicity. Proceedings of the National Academy of Sciences of the United States of America *108*, 16968-16973. 10.1073/pnas.1105270108.

190. Bisker, G., Dong, J., Park, H.D., Iverson, N.M., Ahn, J., Nelson, J.T., Landry, M.P., Kruss, S., and Strano, M.S. (2016). Protein-targeted corona phase molecular recognition. Nature Communications *7*, 10241. 10.1038/ncomms10241.

191. Lazzari, S., Nicoud, L., Jaquet, B., Lattuada, M., and Morbidelli, M. (2016). Fractal-like structures in colloid science. Advances in colloid and interface science *235*, 1-13. 10.1016/j.cis.2016.05.002.

192. Walkey, C., and Chan, W.C.W. (2011). Understanding and controlling the interaction of nanomaterials with proteins in a physiological environment. Chemical Society reviews *41*, 2780-2799. 10.1039/c1cs15233e.

193. Nißler, R., Mann, F.A., Chaturvedi, P., Horlebein, J., Meyer, D., Vuković, L., and Kruss, S. (2019). Quantification of the Number of Adsorbed DNA Molecules on Single-Walled Carbon Nanotubes. The Journal of Physical Chemistry C *123*, 4837-4847. 10.1021/acs.jpcc.8b11058.

194. Safaee, M.M., Gravely, M., Rocchio, C., Simmeth, M., and Roxbury, D. (2019). DNA Sequence Mediates Apparent Length Distribution in Single-Walled Carbon Nanotubes. ACS Applied Materials & Interfaces *11*, 2225-2233. 10.1021/acsami.8b16478.

195. Alizadehmojarad, A.A., Zhou, X., Beyene, A.G., Chacon, K.E., Sung, Y., Pinals, R.L., Landry, M.P., and Vuković, L. (2020). Binding Affinity and Conformational Preferences Influence Kinetic Stability of Short Oligonucleotides on Carbon Nanotubes. Advanced Materials Interfaces *7*, 2000353. 10.1002/admi.202000353.

196. O'Connell, M.J., Bachilo, S.M., Huffman, C.B., Moore, V.C., Strano, M.S., Haroz, E.H., Rialon, K.L., Boul, P.J., Noon, W.H., Kittrell, C., et al. (2002). Band Gap Fluorescence from Individual Single-Walled Carbon Nanotubes. Science *297*, 593-596. doi:10.1126/science.1072631.

197. Jeng, E.S., Moll, A.E., Roy, A.C., Gastala, J.B., and Strano, M.S. (2006). Detection of DNA hybridization using the near-infrared band-gap fluorescence of single-walled carbon nanotubes. Nano letters *6*, 371-375. 10.1021/nl051829k.

198. Schöppler, F., Mann, C., Hain, T.C., Neubauer, F.M., Privitera, G., Bonaccorso, F., Chu, D., Ferrari, A.C., and Hertel, T. (2011). Molar Extinction Coefficient of Single-Wall Carbon Nanotubes. The Journal of Physical Chemistry C *115*, 14682-14686. 10.1021/jp205289h.

199. Antonucci, A., Kupis-Rozmysłowicz, J., and Boghossian, A.A. (2017). Noncovalent Protein and Peptide Functionalization of Single-Walled Carbon Nanotubes for Biodelivery and Optical Sensing Applications. ACS Applied Materials & Interfaces *9*, 11321-11331. 10.1021/acsami.7b00810.

200. Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat Methods *6*, 359-362. 10.1038/nmeth.1322.

201. Simonsen, J.B., and Munter, R. (2020). Pay Attention to Biological Nanoparticles when Studying the Protein Corona on Nanomedicines. Angew Chem Int Ed Engl *59*, 12584-12588. 10.1002/anie.202004611.

202. O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. Journal of Biological Chemistry *250*, 4007-4021. 10.1016/s0021-9258(19)41496-8.

203. Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature *227*, 680-685. 10.1038/227680a0.

204. Hoogland, C., Mostaguir, K., Appel, R.D., and Lisacek, F. (2008). The World-2DPAGE Constellation to promote and publish gel-based proteomics data through the ExPASy server. J Proteomics *71*, 245-248. 10.1016/j.jprot.2008.02.005.

205. Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., and Nicholson, J.K. (2006). UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. Rapid Commun Mass Spectrom *20*, 1989-1994. 10.1002/rcm.2550.

206. Shliaha, P.V., Bond, N.J., Gatto, L., and Lilley, K.S. (2013). Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. J Proteome Res *12*, 2323-2339. 10.1021/pr300775k.

207. Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H., and Tenzer, S. (2014). Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. Nature Methods *11*, 167-170. 10.1038/nmeth.2767.

208. Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics *5*, 144-156. 10.1074/mcp.M500230-MCP200.

209. Neilson, K.A., Ali, N.A., Muralidharan, S., Mirzaei, M., Mariani, M., Assadourian, G., Lee, A., van Sluyter, S.C., and Haynes, P.A. (2011). Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics *11*, 535-553. 10.1002/pmic.201000553.

210. Nahnsen, S., Bielow, C., Reinert, K., and Kohlbacher, O. (2013). Tools for label-free peptide quantification. Mol Cell Proteomics *12*, 549-556. 10.1074/mcp.R112.025163.

211. Thomas, P., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome research *13*, 2129-2141. 10.1101/gr.772403.

212. Bateman, A. (2018). UniProt: A worldwide hub of protein knowledge. Nucleic acids research *47*, D506-D515. 10.1093/nar/gky1049.

213. Findlay, M.R., Freitas, D.N., Mobed-Miremadi, M., and Wheeler, K.E. (2017). Machine learning provides predictive analysis into silver nanoparticle protein corona formation from physicochemical properties. Environmental science. Nano *5*, 64-71. 10.1039/c7en00466d.

214. Schon, A., Clarkson, B.R., Jaime, M., and Freire, E. (2017). Temperature stability of proteins: Analysis of irreversible denaturation using isothermal calorimetry. Proteins *85*, 2009-2016. 10.1002/prot.25354.

215. Kikhney, A.G., and Svergun, D.I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Letters *589*, 2570-2577. 10.1016/j.febslet.2015.08.027.

216. Ackermann, J., Metternich, J.T., Herbertz, S., and Kruss, S. (2022). Biosensing with Fluorescent Carbon Nanotubes. Angewandte Chemie International Edition *61*. 10.1002/anie.202112372.

217. Pinals, R.L., Ledesma, F., Yang, D., Navarro, N., Jeong, S., Pak, J.E., Kuo, L., Chuang, Y.-C., Cheng, Y.-W., Sun, H.-Y., and Landry, M.P. (2021). Rapid SARS-CoV-2 Spike Protein Detection by Carbon Nanotube-Based Near-Infrared Nanosensors. Nano Letters *21*, 2272-2280. 10.1021/acs.nanolett.1c00118.

218. Bates, K., and Kostarelos, K. (2013). Carbon nanotubes as vectors for gene therapy: Past achievements, present challenges and future goals. Advanced Drug Delivery Reviews *65*, 2023-2033. 10.1016/j.addr.2013.10.003.

219. Boyer, P.D., Ganesh, S., Qin, Z., Holt, B.D., Buehler, M.J., Islam, M.F., and Dahl, K.N. (2016). Delivering Single-Walled Carbon Nanotubes to the Nucleus Using Engineered Nuclear Protein Domains. ACS Applied Materials & Interfaces *8*, 3524-3534. 10.1021/acsami.5b12602.

220. Demirer, G.S., Zhang, H., Matos, J.L., Goh, N.S., Cunningham, F.J., Sung, Y., Chang, R., Aditham, A.J., Chio, L., Cho, M.-J., et al. (2019). High aspect ratio nanomaterials enable delivery of functional genetic material without DNA integration in mature plants. Nature Nanotechnology *14*, 456-464. 10.1038/s41565-019-0382-5.

221. Kruss, S., Salem, D.P., Vuković, L., Lima, B., Vander Ende, E., Boyden, E.S., and Strano, M.S. (2017). High-resolution imaging of cellular dopamine efflux using a fluorescent nanosensor array. Proceedings of the National Academy of Sciences *114*, 1789-1794. 10.1073/pnas.1613541114.

222. Bulumulla, C., Krasley, A.T., Cristofori-Armstrong, B., Valinsky, W.C., Walpita, D., Ackerman, D., Clapham, D.E., and Beyene, A.G. (2022). Visualizing synaptic dopamine efflux with a 2D composite nanofilm. eLife *11*, e78773. 10.7554/eLife.78773.

223. Elizarova, S., Chouaib, A.A., Shaib, A., Hill, B., Mann, F., Brose, N., Kruss, S., and Daniel, J.A. (2022). A fluorescent nanosensor paint detects dopamine release at axonal varicosities with high spatiotemporal resolution. Proceedings of the National Academy of Sciences *119*, e2202842119. 10.1073/pnas.2202842119.

224. High-throughput evolution of near-infrared serotonin nanosensors. (2019). SCIENCE ADVANCES.

225. Wu, H., Nißler, R., Morris, V., Herrmann, N., Hu, P., Jeon, S.-J., Kruss, S., and Giraldo, J.P. (2020). Monitoring Plant Health with Near-Infrared Fluorescent H_2O_2 Nanosensors. Nano Letters *20*, 2432-2442. 10.1021/acs.nanolett.9b05159.

226. Lew, T.T.S., Koman, V.B., Silmore, K.S., Seo, J.S., Gordiichuk, P., Kwak, S.-Y., Park, M., Ang, M.C.-Y., Khong, D.T., Lee, M.A., et al. (2020). Real-time detection of wound-induced H2O2

signalling waves in plants with optical nanosensors. Nature Plants *6*, 404-415. 10.1038/s41477-020-0632-4.

227. Safaee, M.M., Gravely, M., and Roxbury, D. (2021). A Wearable Optical Microfibrous Biomaterial with Encapsulated Nanosensors Enables Wireless Monitoring of Oxidative Stress. Advanced Functional Materials *31*, 2006254. 10.1002/adfm.202006254.

228. Zhang, J., Boghossian, A.A., Barone, P.W., Rwei, A., Kim, J.-H., Lin, D., Heller, D.A., Hilmer, A.J., Nair, N., Reuel, N.F., and Strano, M.S. (2011). Single Molecule Detection of Nitric Oxide Enabled by d(AT) $_{15}$ DNA Adsorbed to Near Infrared Fluorescent Single-Walled Carbon Nanotubes. Journal of the American Chemical Society *133*, 567-581. 10.1021/ja1084942.

229. Landry, M.P., Vuković, L., Kruss, S., Bisker, G., Landry, A.M., Islam, S., Jain, R., Schulten, K., and Strano, M.S. (2015). Comparative Dynamics and Sequence Dependence of DNA and RNA Binding to Single Walled Carbon Nanotubes. The Journal of Physical Chemistry C *119*, 10048-10058. 10.1021/jp511448e.

230. Zheng, M., Jagota, A., Semke, E.D., Diner, B.A., Mclean, R.S., Lustig, S.R., Richardson, R.E., and Tassi, N.G. (2003). DNA-assisted dispersion and separation of carbon nanotubes. Nature Materials *2*, 338-342. 10.1038/nmat877.

231. Campbell, J.F., Tessmer, I., Thorp, H.H., and Erie, D.A. (2008). Atomic Force Microscopy Studies of DNA-Wrapped Carbon Nanotube Structure and Binding to Quantum Dots. Journal of the American Chemical Society *130*, 10648-10655. 10.1021/ja801720c.

232. Kwak, S.-Y., Lew, T.T.S., Sweeney, C.J., Koman, V.B., Wong, M.H., Bohmert-Tatarev, K., Snell, K.D., Seo, J.S., Chua, N.-H., and Strano, M.S. (2019). Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers. Nature Nanotechnology *14*, 447-455. 10.1038/s41565-019-0375-4.

233. Hough, L.A., Islam, M.F., Hammouda, B., Yodh, A.G., and Heiney, P.A. (2006). Structure of Semidilute Single-Wall Carbon Nanotube Suspensions and Gels. Nano Letters *6*, 313-317. 10.1021/nl051871f.

234. Barrejón, M., Syrgiannis, Z., Burian, M., Bosi, S., Montini, T., Fornasiero, P., Amenitsch, H., and Prato, M. (2019). Cross-Linked Carbon Nanotube Adsorbents for Water Treatment: Tuning the Sorption Capacity through Chemical Functionalization. ACS Applied Materials & Interfaces *11*, 12920-12930. 10.1021/acsami.8b20557.

235. Pinals, R.L., Yang, D., Rosenberg, D.J., Chaudhary, T., Crothers, A.R., Iavarone, A.T., Hammel, M., and Landry, M.P. (2020). Quantitative Protein Corona Composition and Dynamics on Carbon Nanotubes in Biological Environments. Angewandte Chemie International Edition *59*, 23668-23677. 10.1002/anie.202008175.

236. Demirer, G.S., Zhang, H., Goh, N.S., Pinals, R.L., Chang, R., and Landry, M.P. (2020). Carbon nanocarriers deliver siRNA to intact plant cells for efficient gene knockdown. Science Advances *6*, eaaz0495. 10.1126/sciadv.aaz0495.

237. Roxbury, D., Mittal, J., and Jagota, A. (2012). Molecular-Basis of Single-Walled Carbon Nanotube Recognition by Single-Stranded DNA. Nano Letters *12*, 1464-1469. 10.1021/nl204182b.

238. Yang, D., Yang, S.J., Del Bonis-O'Donnell, J.T., Pinals, R.L., and Landry, M.P. (2020). Mitigation of Carbon Nanotube Neurosensor Induced Transcriptomic and Morphological

Changes in Mouse Microglia with Surface Passivation. ACS Nano *14*, 13794-13805. 10.1021/acsnano.0c06154.

239.   Dulkeith, E., Ringler, M., Klar, T.A., Feldmann, J., Muñoz Javier, A., and Parak, W.J. (2005). Gold Nanoparticles Quench Fluorescence by Phase Induced Radiative Rate Suppression. Nano Letters *5*, 585-589. 10.1021/nl0480969.

240.   Zhu, Z., Yang, R., You, M., Zhang, X., Wu, Y., and Tan, W. (2010). Single-walled carbon nanotube as an effective quencher. Analytical and Bioanalytical Chemistry *396*, 73-83. 10.1007/s00216-009-3192-z.

241.   Brege, J.J., Gallaway, C., and Barron, A.R. (2009). Fluorescence Quenching of Single-Walled Carbon Nanotubes with Transition-Metal Ions. The Journal of Physical Chemistry C *113*, 4270-4276. 10.1021/jp808667b.

242.   A, B., G, F., and O, G. (2000). Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). Journal of Applied Crystallography *33*, 1212-1216. 10.1107/S0021889800008372.

243.   Putnam, C.D.H., Michal; Hura, Greg L.; Tainer, John A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly reviews of biophysics *40*, 191-285. 10.1017/s0033583507004635.

244.   Bisker, G., Ahn, J., Kruss, S., Ulissi, Z.W., Salem, D.P., and Strano, M.S. (2015). A Mathematical Formulation and Solution of the CoPhMoRe Inverse Problem for Helically Wrapping Polymer Corona Phases on Cylindrical Substrates. The Journal of Physical Chemistry C *119*, 13876-13886. 10.1021/acs.jpcc.5b01705.

245.   Johnson, R.R., Johnson, A.T.C., and Klein, M.L. (2008). Probing the Structure of DNA−Carbon Nanotube Hybrids with Molecular Dynamics. Nano Letters *8*, 69-75. 10.1021/nl071909j.

246.   Zheng, M., Jagota, A., Strano, M.S., Santos, A.P., Barone, P., Chou, S.G., Diner, B.A., Dresselhaus, M.S., Mclean, R.S., Onoa, G.B., et al. (2003). Structure-Based Carbon Nanotube Sorting by Sequence-Dependent DNA Assembly. Science *302*, 1545-1548. 10.1126/science.1091911.

247.   Cathcart, H., Nicolosi, V., Hughes, J.M., Blau, W.J., Kelly, J.M., Quinn, S.J., and Coleman, J.N. (2008). Ordered DNA Wrapping Switches on Luminescence in Single-Walled Nanotube Dispersions. Journal of the American Chemical Society *130*, 12734-12744. 10.1021/ja803273s.

248.   Jin, H., Jeng, E.S., Heller, D.A., Jena, P.V., Kirmse, R., Langowski, J., and Strano, M.S. (2007). Divalent Ion and Thermally Induced DNA Conformational Polymorphism on Single-walled Carbon Nanotubes. Macromolecules *40*, 6731-6739. 10.1021/ma070608t.

249.   Salem, D.P., Gong, X., Liu, A.T., Koman, V.B., Dong, J., and Strano, M.S. (2017). Ionic Strength-Mediated Phase Transitions of Surface-Adsorbed DNA on Single-Walled Carbon Nanotubes. Journal of the American Chemical Society *139*, 16791-16802. 10.1021/jacs.7b09258.

250.   Gillen, A.J., Kupis-Rozmysłowicz, J., Gigli, C., Schuergers, N., and Boghossian, A.A. (2018). Xeno Nucleic Acid Nanosensors for Enhanced Stability Against Ion-Induced Perturbations. The Journal of Physical Chemistry Letters *9*, 4336-4343. 10.1021/acs.jpclett.8b01879.

251.   Chen, K.-I., Li, B.-R., and Chen, Y.-T. (2011). Silicon nanowire field-effect transistor-based biosensors for biomedical diagnosis and cellular recording investigation. Nano Today *6*, 131-154. 10.1016/j.nantod.2011.02.001.

252. Ghosh, S., Patel, N., and Chakrabarti, R. (2016). Probing the Salt Concentration Dependent Nucleobase Distribution in a Single-Stranded DNA–Single-Walled Carbon Nanotube Hybrid with Molecular Dynamics. The Journal of Physical Chemistry B *120*, 455-466. 10.1021/acs.jpcb.5b12044.

253. Ivchenko, E.L., Kochereshko, V.P., Kop'ev, P.S., Kosobukin, V.A., Uraltsev, I.N., and Yakovlev, D.R. EXCITON LONGITUDINAL-TRANSVERSE SPLITTING IN GaAs/AIGaAs SUPERLATTICES AND MULTIPLE QUANTUM W]~Y,T.q. *70*.

254. Orthaber, D., Bergmann, A., and Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. Journal of Applied Crystallography *33*, 218-225. 10.1107/S0021889899015216.

255. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676-682. 10.1038/nmeth.2019.

256. Burian, M., and Amenitsch, H. (2018). Dummy-atom modelling of stacked and helical nanostructures from solution scattering data. IUCrJ *5*, 390-401. 10.1107/S2052252518005493.

257. Peigney, A., Laurent, C., Flahaut, E., Bacsa, R.R., and Rousset, A. (2001). Specific surface area of carbon nanotubes and bundles of carbon nanotubes. Carbon *39*, 507-514. 10.1016/S0008-6223(00)00155-X.

258. Mori, T., and Hegmann, T. (2016). Determining the composition of gold nanoparticles: a compilation of shapes, sizes, and calculations using geometric considerations. Journal of Nanoparticle Research *18*, 295. 10.1007/s11051-016-3587-7.

259. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry *25*, 1605-1612. 10.1002/jcc.20084.

260. Bittova, B., Vejpravova, J.P., Kalbac, M., Burianova, S., Mantlikova, A., Danis, S., and Doyle, S. (2011). Magnetic Properties of Iron Catalyst Particles in HiPco Single Wall Carbon Nanotubes. The Journal of Physical Chemistry C *115*, 17303-17309. 10.1021/jp203365g.

261. Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal *76*, 2879-2886. 10.1016/s0006-3495(99)77443-6.

262. Svergun, D.I., Petoukhov, M.V., and Koch, M.H.J. (2001). Determination of Domain Structure of Proteins from X-Ray Solution Scattering. Biophysical Journal *80*, 2946-2953. 10.1016/S0006-3495(01)76260-1.

263. Grant, T.D. (2018). Ab initio electron density determination directly from solution scattering data. Nature Methods *15*, 191-193. 10.1038/nmeth.4581.

264. Sharif, H., Li, Y., Dong, Y., Dong, L., Wang, W.L., Mao, Y., and Wu, H. (2017). Cryo-EM structure of the DNA-PK holoenzyme. Proc Natl Acad Sci U S A *114*, 7367-7372. 10.1073/pnas.1707386114.

265. Baretic, D., Maia de Oliveira, T., Niess, M., Wan, P., Pollard, H., Johnson, C.M., Truman, C., McCall, E., Fisher, D., Williams, R., and Phillips, C. (2019). Structural insights into the critical DNA damage sensors DNA-PKcs, ATM and ATR. Progress in biophysics and molecular biology *147*, 4-16. 10.1016/j.pbiomolbio.2019.06.003.

266. Sibanda, B.L., Chirgadze, D.Y., Ascher, D.B., and Blundell, T.L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. Science *355*, 520-524. 10.1126/science.aak9654.

267. Dutta, A., Eckelmann, B., Adhikari, S., Ahmed, K.M., Sengupta, S., Pandey, A., Hegde, P.M., Tsai, M.S., Tainer, J.A., Weinfeld, M., et al. (2017). Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex. Nucleic acids research *45*, 2585-2599. 10.1093/nar/gkw1262.

268. Lafrance-Vanasse, J., Williams, G.J., and Tainer, J.A. (2015). Envisioning the dynamics and flexibility of Mre11-Rad50-Nbs1 complex to decipher its roles in DNA replication and repair. Progress in biophysics and molecular biology *117*, 182-193. 10.1016/j.pbiomolbio.2014.12.004.

269. Syed, A., and Tainer, J.A. (2018). The MRE11-RAD50-NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. Annu Rev Biochem *87*, 263-294. 10.1146/annurev-biochem-062917-012415.

270. Williams, G.J., Hammel, M., Radhakrishnan, S.K., Ramsden, D., Lees-Miller, S.P., and Tainer, J.A. (2014). Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. DNA Repair (Amst) *17*, 110-120. 10.1016/j.dnarep.2014.02.009.

271. Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson, E.A., Giovannangeli, C., et al. (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Molecular cell *55*, 829-842. 10.1016/j.molcel.2014.08.002.

272. Bacolla, A., Tainer, J.A., Vasquez, K.M., and Cooper, D.N. (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic acids research. 10.1093/nar/gkw261.

273. Spagnolo, L., Barbeau, J., Curtin, N.J., Morris, E.P., and Pearl, L.H. (2012). Visualization of a DNA-PK/PARP1 complex. Nucleic acids research *40*, 4168-4177. 10.1093/nar/gkr1231.

274. Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature *434*, 913-917. 10.1038/nature03443.

275. Walker, J.R., Corpina, R.A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature *412*, 607-614.

276. Yin, X., Liu, M., Tian, Y., Wang, J., and Xu, Y. (2017). Cryo-EM structure of human DNA-PK holoenzyme. Cell research *27*, 1341-1350. 10.1038/cr.2017.110.

277. Hammel, M. (2012). Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). Eur Biophys J *41*, 789-799.

278. Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quarterly reviews of biophysics *40*, 191-285.

279. Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Current opinion in structural biology *58*, 197-213. 10.1016/j.sbi.2019.04.004.

280.   Hura, G.L., Budworth, H., Dyer, K.N., Rambo, R.P., Hammel, M., McMurray, C.T., and Tainer, J.A. (2013). Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. Nature methods. nmeth.2453 [pii] 10.1038/nmeth.2453.

281.   Rambo, R.P., and Tainer, J.A. (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers *95*, 559-571. 10.1002/bip.21638.

282.   Schneidman-Duhovny, D., and Hammel, M. (2018). Modeling Structure and Dynamics of Protein Complexes with SAXS Profiles. Methods Mol Biol *1764*, 449-473. 10.1007/978-1-4939-7759-8_29.

283.   Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen Physiol Biophys *28*, 174-189.

284.   Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic acids research *44*, W424-429. 10.1093/nar/gkw389.

285.   Horst, B.G., Yokom, A.L., Rosenberg, D.J., Morris, K.L., Hammel, M., Hurley, J.H., and Marletta, M.A. (2019). Allosteric activation of the nitric oxide receptor soluble guanylate cyclase mapped by cryo-electron microscopy. Elife *8*. 10.7554/eLife.50634.

286.   Hammel, M., Yu, Y., Mahaney, B.L., Cai, B., Ye, R., Phipps, B.M., Rambo, R.P., Hura, G.L., Pelikan, M., So, S., et al. (2010). Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. J Biol Chem *285*, 1414-1423. M109.065615 [pii] 10.1074/jbc.M109.065615.

287.   Zhou, Y., Millott, R., Kim, H.J., Peng, S., Edwards, R.A., Skene-Arnold, T., Hammel, M., Lees-Miller, S.P., Tainer, J.A., Holmes, C.F.B., and Glover, J.N.M. (2019). Flexible Tethering of ASPP Proteins Facilitates PP-1c Catalysis. Structure *27*, 1485-1496 e1484. 10.1016/j.str.2019.07.012.

288.   Hammel, M., Yu, Y., Radhakrishnan, S.K., Chokshi, C., Tsai, M.S., Matsumoto, Y., Kuzdovich, M., Remesh, S.G., Fang, S., Tomkinson, A.E., et al. (2016). An Intrinsically Disordered APLF Links Ku, DNA-PKcs and XRCC4-DNA Ligase IV in an Extended Flexible Non-Homologous End Joining Complex. J Biol Chem. 10.1074/jbc.M116.751867.

289.   Panjkovich, A., and Svergun, D.I. (2016). Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. Phys Chem Chem Phys *18*, 5707-5719. 10.1039/c5cp04540a.

290.   Kashammer, L., Saathoff, J.H., Lammens, K., Gut, F., Bartho, J., Alt, A., Kessler, B., and Hopfner, K.P. (2019). Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. Molecular cell *76*, 382-394 e386. 10.1016/j.molcel.2019.07.035.

291.   Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys J *105*, 962-974. 10.1016/j.bpj.2013.07.020.

292.   Rivera-Calzada, A., Spagnolo, L., Pearl, L.H., and Llorca, O. (2007). Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. EMBO reports *8*, 56-62.

293. Andrews, B.J., Lehman, J.A., and Turchi, J.J. (2006). Kinetic analysis of the Ku-DNA binding activity reveals a redox-dependent alteration in protein structure that stimulates dissociation of the Ku-DNA complex. J Biol Chem *281*, 13596-13603. 10.1074/jbc.M512787200.

294. Hartley, K.O., Gell, D., Smith, G.C., Zhang, H., Divecha, N., Connelly, M.A., Admon, A., Lees-Miller, S.P., Anderson, C.W., and Jackson, S.P. (1995). DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. Cell *82*, 849-856. 10.1016/0092-8674(95)90482-4.

295. Williams, D.R., Lee, K.J., Shi, J., Chen, D.J., and Stewart, P.L. (2008). Cryo-EM Structure of the DNA-Dependent Protein Kinase Catalytic Subunit at Subnanometer Resolution Reveals alpha Helices and Insight into DNA Binding. Structure *16*, 468-477.

296. Rivera-Calzada, A., Maman, J.D., Spagnolo, L., Pearl, L.H., and Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). Structure *13*, 243-255.

297. Sibanda, B.L., Chirgadze, D.Y., and Blundell, T.L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature *463*, 118-121. nature08648 [pii] 10.1038/nature08648.

298. Douglas, P., Sapkota, G.P., Morrice, N., Yu, Y., Goodarzi, A.A., Merkle, D., Meek, K., Alessi, D.R., and Lees-Miller, S.P. (2002). Identification of in vitro and in vivo phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. The Biochemical journal *368*, 243-251.

299. Ding, Q., Reddy, Y.V., Wang, W., Woods, T., Douglas, P., Ramsden, D.A., Lees-Miller, S.P., and Meek, K. (2003). Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase Is required for efficient end processing during DNA double-strand break repair. Molecular and cellular biology *23*, 5836-5848.

300. Lees-Miller, J.P., Cobban, A., Katsonis, P., Bacolla, A., Tsutakawa, S.E., Hammel, M., Meek, K., Anderson, D.W., Lichtarge, O., Tainer, J.A., and Lees-Miller, S.P. (2020). Uncovering DNA-PKcs ancient phylogeny, unique sequence motifs and insights for human disease. Progress in biophysics and molecular biology *this issue*.

301. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology *234*, 779-815. S0022-2836(83)71626-8 [pii] 10.1006/jmbi.1993.1626.

302. Douglas, P., Cui, X., Block, W.D., Yu, Y., Gupta, S., Ding, Q., Ye, R., Morrice, N., Lees-Miller, S.P., and Meek, K. (2007). The DNA-dependent protein kinase catalytic subunit is phosphorylated in vivo on threonine 3950, a highly conserved amino acid in the protein kinase domain. Molecular and cellular biology *27*, 1581-1591.

303. Block, W.D., Yu, Y., Merkle, D., Gifford, J.L., Ding, Q., Meek, K., and Lees-Miller, S.P. (2004). Autophosphorylation-dependent remodeling of the DNA-dependent protein kinase catalytic subunit regulates ligation of DNA ends. Nucleic acids research *32*, 4351-4357.

304. Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P.O., van Gent, D.C., Chen, B.P., and Chen, D.J. (2007). Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. The Journal of cell biology *177*, 219-229.

305.   Jette, N., and Lees-Miller, S.P. (2014). The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. Progress in biophysics and molecular biology. 10.1016/j.pbiomolbio.2014.12.003.

306.   Spagnolo, L., Rivera-Calzada, A., Pearl, L.H., and Llorca, O. (2006). Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. Molecular cell *22*, 511-519.

307.   Yoo, S., and Dynan, W.S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. Nucleic acids research *27*, 4679-4686.

308.   Boskovic, J., Rivera-Calzada, A., Maman, J.D., Chacon, P., Willison, K.R., Pearl, L.H., and Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. The EMBO journal *22*, 5875-5882.

309.   Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic acids research *33*, W363-367. 33/suppl_2/W363 [pii] 10.1093/nar/gki481.

310.   Yang, H., Jiang, X., Li, B., Yang, H.J., Miller, M., Yang, A., Dhar, A., and Pavletich, N.P. (2017). Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. Nature *552*, 368-373. 10.1038/nature25023.

311.   Rambo, R.P., and Tainer, J.A. (2013). Super-Resolution in Solution X-Ray Scattering and Its Applications to Structural Systems Biology. Annu Rev Biophys. 10.1146/annurev-biophys-083012-130301.

312.   Yano, K., and Chen, D.J. (2008). Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. Cell cycle (Georgetown, Tex *7*, 1321-1325.

313.   Cottarel, J., Frit, P., Bombarde, O., Salles, B., Negrel, A., Bernard, S., Jeggo, P.A., Lieber, M.R., Modesti, M., and Calsou, P. (2013). A noncatalytic function of the ligation complex during nonhomologous end joining. The Journal of cell biology *200*, 173-186. 10.1083/jcb.201203128.

314.   Wang, C., and Lees-Miller, S.P. (2013). Detection and repair of ionizing radiation-induced DNA double strand breaks: new developments in nonhomologous end joining. International journal of radiation oncology, biology, physics *86*, 440-449. 10.1016/j.ijrobp.2013.01.011.

315.   Hammel, M., Rey, M., Yu, Y., Mani, R.S., Classen, S., Liu, M., Pique, M.E., Fang, S., Mahaney, B.L., Weinfeld, M., et al. (2011). XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair. J Biol Chem *286*, 32638-32650. M111.272641 [pii] 10.1074/jbc.M111.272641.

316.   Hammel, M., Yu, Y., Fang, S., Lees-Miller, S.P., and Tainer, J.A. (2010). XLF regulates filament architecture of the XRCC4.ligase IV complex. Structure *18*, 1431-1442. S0969-2126(10)00358-8 [pii] 10.1016/j.str.2010.09.009.

317.   Wang, J.L., Duboc, C., Wu, Q., Ochi, T., Liang, S., Tsutakawa, S.E., Lees-Miller, S.P., Nadal, M., Tainer, J.A., Blundell, T.L., and Strick, T.R. (2018). Dissection of DNA double-strand-break repair using novel single-molecule forceps. Nat Struct Mol Biol *25*, 482-487. 10.1038/s41594-018-0065-1.

318.   Reid, D.A., Keegan, S., Leo-Macias, A., Watanabe, G., Strande, N.T., Chang, H.H., Oksuz, B.A., Fenyo, D., Lieber, M.R., Ramsden, D.A., and Rothenberg, E. (2015). Organization and dynamics

of the nonhomologous end-joining machinery during DNA double-strand break repair. Proc Natl Acad Sci U S A. 10.1073/pnas.1420115112.

319.   Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, L.D., Jardine, E., Gotham, V.J.B., et al. (2017). Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. Nature communications *8*, 15855. 10.1038/ncomms15855.

320.   Rashid, F., Harris, P.D., Zaher, M.S., Sobhy, M.A., Joudeh, L.I., Yan, C., Piwonski, H., Tsutakawa, S.E., Ivanov, I., Tainer, J.A., et al. (2017). Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. Elife *6*. 10.7554/eLife.21884.

321.   Paillard, S., and Strauss, F. (1993). Site-specific proteolytic cleavage of Ku protein bound to DNA. Proteins *15*, 330-337.

322.   Mo, X., and Dynan, W.S. (2002). Subnuclear localization of Ku protein: functional association with RNA polymerase II elongation sites. Molecular and cellular biology *22*, 8088-8099.

323.   Harris, R., Esposito, D., Sankar, A., Maman, J.D., Hinks, J.A., Pearl, L.H., and Driscoll, P.C. (2004). The 3D solution structure of the C-terminal region of Ku86 (Ku86CTR). Journal of molecular biology *335*, 573-582.

324.   Gell, D., and Jackson, S.P. (1999). Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. Nucleic acids research *27*, 3494-3502.

325.   Chen, B.P., Chan, D.W., Kobayashi, J., Burma, S., Asaithamby, A., Morotomi-Yano, K., Botvinick, E., Qin, J., and Chen, D.J. (2005). Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. J Biol Chem *280*, 14709-14715.

326.   Baretic, D., Pollard, H.K., Fisher, D.I., Johnson, C.M., Santhanam, B., Truman, C.M., Kouba, T., Fersht, A.R., Phillips, C., and Williams, R.L. (2017). Structures of closed and open conformations of dimeric human ATM. Sci Adv *3*, e1700933. 10.1126/sciadv.1700933.

327.   Cui, X., Yu, Y., Gupta, S., Cho, Y.M., Lees-Miller, S.P., and Meek, K. (2005). Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. Molecular and cellular biology *25*, 10842-10852.

328.   Lees-Miller, S.P., Beattie, T.L., and Tainer, J.A. (2016). Noncoding RNA joins Ku and DNA-PKcs for DNA-break resistance in breast cancer. Nat Struct Mol Biol *23*, 509-510. 10.1038/nsmb.3240.

329.   Goodarzi, A.A., and Lees-Miller, S.P. (2004). Biochemical characterization of the ataxia-telangiectasia mutated (ATM) protein from human cells. DNA Repair (Amst) *3*, 753-767.

330.   Chan, D.W., Mody, C.H., Ting, N.S., and Lees-Miller, S.P. (1996). Purification and characterization of the double-stranded DNA-activated protein kinase, DNA-PK, from human placenta. Biochem Cell Biol *74*, 67-73.

331.   Radhakrishnan, S.K., and Lees-Miller, S.P. (2017). DNA requirements for interaction of the C-terminal region of Ku80 with the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). DNA repair *57*, 17-28. 10.1016/j.dnarep.2017.06.001.

332.   Classen, S., Hura, G.L., Holton, J.M., Rambo, R.P., Rodic, I., McGuire, P.J., Dyer, K., Hammel, M., Meigs, G., Frankel, K.A., and Tainer, J.A. (2013). Implementation and performance of

SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. J Appl Crystallogr *46*, 1-13. 10.1107/S0021889812048698.

333. Dyer, K.N., Hammel, M., Rambo, R.P., Tsutakawa, S.E., Rodic, I., Classen, S., Tainer, J.A., and Hura, G.L. (2014). High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. Methods Mol Biol *1091*, 245-258. 10.1007/978-1-62703-691-7_18.

334. Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. Nature *496*, 477-481. 10.1038/nature12070.

335. Rambo, R.P., and Tainer, J.A. (2013). Super-resolution in solution X-ray scattering and its applications to structural systems biology. Annu Rev Biophys *42*, 415-441. 10.1146/annurev-biophys-083012-130301.

336. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic acids research *38*, W540-544. gkq461 [pii] 10.1093/nar/gkq461.

337. Yip, S.H.-C.F., Jee Loon; Schenk, Gerhard; Gahan, Lawrence R.; Carr, Paul D.; Ollis, David L. (2011). Directed evolution combined with rational design increases activity of GpdQ toward a non-physiological substrate and alters the oligomeric structure of the enzyme. Protein engineering, design & selection : PEDS *24*, 861-872. 10.1093/protein/gzr048.

338. Nguyen, T.T.G., Rodolfo; Roche, Julien; Venditti, Vincenzo (2021). Structure elucidation of the elusive Enzyme I monomer reveals the molecular mechanisms linking oligomerization and enzymatic activity. Proceedings of the National Academy of Sciences of the United States of America *118*, NA-NA. 10.1073/pnas.2100298118.

339. Gray, M.W.L., Julius; Archibald, John M.; Keeling, Patrick J.; Doolittle, W. F. (2010). Cell biology. Irremediable complexity. Science (New York, N.Y.) *330*, 920-921. 10.1126/science.1198594.

340. Lynch, M. (2013). Evolutionary diversification of the multimeric states of proteins. Proceedings of the National Academy of Sciences of the United States of America *110*, 201310980-201310988. 10.1073/pnas.1310980110.

341. Hochberg, G.K.A.L., Yang; Marklund, Erik G.; Metzger, Brian P. H.; Laganowsky, Arthur; Thornton, Joseph W. (2020). A hydrophobic ratchet entrenches molecular complexes. Nature *588*, 503-508. 10.1038/s41586-020-3021-2.

342. Schneider, G.L., Y; Brändén, C I (1992). RUBISCO: structure and mechanism. Annual review of biophysics and biomolecular structure *21*, 119-143. 10.1146/annurev.bb.21.060192.001003.

343. Bar-On, Y.M.M., Ron (2019). The global mass and average rate of rubisco. Proceedings of the National Academy of Sciences of the United States of America *116*, 4738-4743. 10.1073/pnas.1816654116.

344. Banda, D.M.P., Jose Henrique; Liu, Albert K; Orr, Douglas J; Hammel, Michal; He, Christine; Parry, Martin A. J.; Carmo-Silva, Elizabete; Adams, Paul D.; Banfield, Jillian F.; Shih, Patrick M. (2020). Novel bacterial clade reveals origin of form I Rubisco. Nature plants *6*, 1158-1166. 10.1038/s41477-020-00762-4.

345. Schneider, G.L., Ylva; Lundqvist, Tomas (1990). Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from Rhodospirillum rubrum at 1.7 Å resolution. Journal of molecular biology *211*, 989-1008. 10.1016/0022-2836(90)90088-4.

346. Satagopan, S.C., Sum; Perry, L. Jeanne; Tabita, F. Robert (2014). Structure-Function Studies with the Unique Hexameric Form II Ribulose-1,5-bisphosphate Carboxylase/Oxygenase (Rubisco) from Rhodopseudomonas palustris*. The Journal of biological chemistry *289*, 21433-21450. 10.1074/jbc.m114.578625.

347. Varaljay, V.A.S., Sriram; North, Justin A.; Witte, Brian; Dourado, Manuella Nóbrega; Anantharaman, Karthik; Arbing, Mark A.; McCann, Shelley Hoeft; Oremland, Ronald S.; Banfield, Jillian F.; Wrighton, Kelly C.; Tabita, F. Robert (2016). Functional metagenomic selection of ribulose 1, 5-bisphosphate carboxylase/oxygenase from uncultivated bacteria. Environmental microbiology *18*, 1187-1199. 10.1111/1462-2920.13138.

348. Davidi, D.S., Melina; Guo, Zhijun; Bar-On, Yinon M.; Prywes, Noam; Oz, Aia; Jabłońska, Jagoda; Flamholz, Avi I.; Wernick, David G.; Antonovsky, Niv; de Pins, Benoit; Shachar, Lior; Hochhauser, Dina; Peleg, Yoav; Albeck, Shira; Sharon, Itai; Mueller-Cajar, Oliver; Milo, Ron (2020). Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. The EMBO journal *39*, e104081-NA. 10.15252/embj.2019104081.

349. Levy, E.D.E., Elisabetta Boeri; Robinson, Carol V.; Teichmann, Sarah A. (2008). Assembly reflects evolution of protein complexes. Nature *453*, 1262-1265. 10.1038/nature06942.

350. Perica, T.C., Cyrus; Teichmann, Sarah A. (2012). Evolution of oligomeric state through geometric coupling of protein interfaces. Proceedings of the National Academy of Sciences of the United States of America *109*, 8127-8132. 10.1073/pnas.1120028109.

351. Kayikci, M.V., AJ; Scott-Brown, James; Ravarani, Charles N. J.; Flock, Tilman; Babu, M. Madan (2018). Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. Nature structural & molecular biology *25*, 185-194. 10.1038/s41594-017-0019-z.

352. Archibald, J.M.L., John M.; Doolittle, W. Ford (1999). Recurrent paralogy in the evolution of archaeal chaperonins. Current biology : CB *9*, 1053-1056. 10.1016/s0960-9822(99)80457-6.

353. Doolittle, W.F. (2012). Evolutionary biology: A ratchet for protein complexity. Nature *481*, 270-271. 10.1038/nature10816.

354. Wang, L.T., Scott; Selzer, Tzvia; Benkovic, Stephen J.; Kohen, Amnon (2006). Effects of a distal mutation on active site chemistry. Biochemistry *45*, 1383-1392. 10.1021/bi0518242.

355. Buller, A.R.B.-C., Sabine; Romney, David K.; Herger, Michael; Murciano-Calles, Javier; Arnold, Frances H. (2015). Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. Proceedings of the National Academy of Sciences of the United States of America *112*, 14599-14604. 10.1073/pnas.1516401112.

356. King, N.P.S., William; Sawaya, Michael R.; Vollmar, Breanna S.; Sumida, John P.; André, Ingemar; Gonen, Tamir; Yeates, Todd O.; Baker, David (2012). Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science (New York, N.Y.) *336*, 1171-1174. 10.1126/science.1219364.

357. Wetzel, C.P., Simone; Picci, Cristina; Gök, Caglar; Hoffmann, Diana; Bali, Kiran Kumar; Lampe, André; Lapatsina, Liudmila; Fleischer, Raluca; St. John Smith, Ewan; Bégay, Valérie; Moroni, Mirko; Estebanez, Luc; Kühnemund, Johannes; Walcher, Jan; Specker, Edgar; Neuenschwander, Martin; von Kries, Jens Peter; Haucke, Volker; Kuner, Rohini; Poulet, James F.A.; Schmoranzer, Jan; Poole, Kate; Lewin, Gary R. (2016). Small-molecule inhibition of STOML3 oligomerization reverses pathological mechanical hypersensitivity. Nature neuroscience *20*, 209-218. 10.1038/nn.4454.

358. Galzitskaya, O.V. (2020). Oligomers Are Promising Targets for Drug Development in the Treatment of Proteinopathies. Frontiers in molecular neuroscience *12*, 319-319. 10.3389/fnmol.2019.00319.

359. Schuster-Böckler, B.B., Alex (2008). Protein interactions in human genetic diseases. Genome biology *9*, 1-12. 10.1186/gb-2008-9-1-r9.

360. Choi, M.L.G., Sonia (2018). Crucial role of protein oligomerization in the pathogenesis of Alzheimer's and Parkinson's diseases. The FEBS journal *285*, 3631-3644. 10.1111/febs.14587.

361. Garcia-Seisdedos, H.E.-M., Charly; Elad, Nadav; Levy, Emmanuel D. (2017). Proteins evolve on the edge of supramolecular self-assembly. Nature *548*, 244-247. 10.1038/nature23320.

362. Hashimoto, K.P., Anna R. (2010). Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. Proceedings of the National Academy of Sciences of the United States of America *107*, 20352-20357. 10.1073/pnas.1012999107.

363. Hashimoto, K.N., Hafumi; Bryant, Stephen H.; Panchenko, Anna R. (2011). Caught in self-interaction: evolutionary and functional mechanisms of protein homooligomerization. Physical biology *8*, 035007-035007. 10.1088/1478-3975/8/3/035007.

364. Goodsell, D.S.O., Arthur J. (2000). Structural symmetry and protein function. Annual review of biophysics and biomolecular structure *29*, 105-153. 10.1146/annurev.biophys.29.1.105.

365. Andrews, T.J. (1988). Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. The Journal of biological chemistry *263*, 12213-12219. 10.1016/s0021-9258(18)37741-x.

366. Li, W.G., Adam (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England) *22*, 1658-1659. 10.1093/bioinformatics/btl158.

367. Katoh, K.R., John; Yamada, Kazunori D. (2017). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics *20*, 1160-1166. 10.1093/bib/bbx108.

368. Darriba, D.T., Guillermo L.; Doallo, Ramón; Posada, David (2011). ProtTest 3. Bioinformatics (Oxford, England) *27*, 1164-1165. 10.1093/bioinformatics/btr088.

369. Mayrose, I.G., Dan; Ben-Tal, Nir; Pupko, Tal (2004). Comparison of Site-Specific Rate-Inference Methods for Protein Sequences: Empirical Bayesian Methods Are Superior. Molecular biology and evolution *21*, 1781-1791. 10.1093/molbev/msh194.

370. Frey, S.G., Dirk (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. Journal of chromatography. A *1337*, 95-105. 10.1016/j.chroma.2014.02.029.

371. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Research *38*, W540-W544. 10.1093/nar/gkq461.

372. Schneidman-Duhovny, D., Hammel, M., Tainer, John A., and Sali, A. (2013). Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. Biophysical Journal *105*, 962-974. 10.1016/j.bpj.2013.07.020.

373. Alonso, H.B., Michelle J.; Beck, Jennifer L.; Whitney, Spencer M. (2009). Substrate-induced assembly of Methanococcoides burtonii D-ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. The Journal of biological chemistry *284*, 33876-33882. 10.1074/jbc.m109.050989.

374. Pereira, J.H.M., Ryan P.; Tomaleri, Giovani P.; Adams, Paul D. (2017). Berkeley Screen: a set of 96 solutions for general macromolecular crystallization. Journal of applied crystallography *50*, 1352-1358. 10.1107/s1600576717011347.

375. Winter, G.L., C.M.C.; Prince, Stephen M. (2013). Decision making in xia2. Acta crystallographica. Section D, Biological crystallography *69*, 1260-1273. 10.1107/s0907444913015308.

376. McCoy, A.J.G.-K., Ralf W.; Adams, Paul D.; Winn, Martyn; Storoni, Laurent C.; Read, Randy J. (2007). Phaser crystallographic software. Journal of applied crystallography *40*, 658-674. 10.1107/s0021889807021206.

377. Terwilliger, T.C.G.-K., Ralf W.; Afonine, Pavel V.; Moriarty, Nigel W.; Zwart, Peter H.; Hung, Li-Wei; Read, Randy J.; Adams, Paul D. (2007). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta crystallographica. Section D, Biological crystallography *64*, 61-69. 10.1107/s090744490705024x.

378. Adams, P.D.A., Pavel V.; Bunkóczi, Gábor; Chen, Vincent B.; Davis, Ian W.; Echols, Nathaniel; Headd, Jeffrey J.; Hung, Li-Wei; Kapral, Gary J.; Grosse-Kunstleve, Ralf W.; McCoy, Airlie J.; Moriarty, Nigel W.; Oeffner, Robert D.; Read, Randy J.; Richardson, David S.; Richardson, Jane S.; Terwilliger, Thomas C.; Zwart, Peter H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography *66*, 213-221. 10.1107/s0907444909052925.

379. Afonine, P.V.G.-K., Ralf W.; Echols, Nathaniel; Headd, Jeffrey J.; Moriarty, Nigel W.; Mustyakimov, Marat; Terwilliger, Thomas C.; Urzhumtsev, Alexandre; Zwart, Peter H.; Adams, Paul D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica. Section D, Biological crystallography *68*, 352-367. 10.1107/s0907444912001308.

380. Emsley, P.C., Kevin (2004). Coot: model-building tools for molecular graphics. Acta crystallographica. Section D, Biological crystallography *60*, 2126-2132. 10.1107/s0907444904019158.

381. Davis, I.W.L.-F., Andrew; Chen, Vincent B.; Block, Jeremy N.; Kapral, Gary J.; Wang, Xueyi; Murray, Laura Weston; Arendall, W. Bryan; Snoeyink, Jack; Richardson, Jane S.; Richardson, David S. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic acids research *35*, 375-383. 10.1093/nar/gkm216.

382. Prins, A.O., Douglas J.; Andralojc, P. John; Reynolds, Matthew P.; Carmo-Silva, Elizabete; Parry, Martin A. J. (2016). Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. Journal of experimental botany *67*, 1827-1838. 10.1093/jxb/erv574.

383. Sharwood, R.E.G., Oula; Whitney, Spencer M. (2016). Prospects for improving CO_2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity. Current opinion in plant biology *31*, 135-142. 10.1016/j.pbi.2016.04.002.

384. Parry, M.A.J.K., Alfred J.; Gutteridge, Steven (1989). Variation in the Specificity Factor of C3 Higher Plant Rubiscos Determined by the Total Consumption of Ribulose-P2. Journal of Experimental Botany *40*, 317-320. 10.1093/jxb/40.3.317.

385. Song, Y.D., Frank; Wang, Ray Yu-Ruei; Kim, David E.; Miles, Chris; Brunette, T. J.; Thompson, James; Baker, David (2013). High resolution comparative modeling with RosettaCM. Structure (London, England : 1993) *21*, 1735-1742. 10.1016/j.str.2013.08.005.

386. Madeira, F.P., Youngmi; Lee, Joon; Buso, Nicola; Gur, Tamer; Madhusoodanan, Nandana; Basutkar, Prasad; Tivey, Adrian R N; Potter, Simon C.; Finn, Robert D.; Lopez, Rodrigo (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research *47*, W636-W641. 10.1093/nar/gkz268.

387. Shapovalov, M.V.D., Roland L. (2011). A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure (London, England : 1993) *19*, 844-858. 10.1016/j.str.2011.03.019.

388. Chaudhury, S.L., Sergey; Gray, Jeffrey J. (2010). PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics (Oxford, England) *26*, 689-691. 10.1093/bioinformatics/btq007.

389. Khatib, F.C., Seth; Tyka, Michael D.; Xu, Kefan; Makedon, Ilya; Popović, Zoran; Baker, David; Players, Foldit (2011). Algorithm discovery by protein folding game players. Proceedings of the National Academy of Sciences of the United States of America *108*, 18949-18953. 10.1073/pnas.1115898108.

390. Alford, R.F.L.-F., Andrew; Jeliazkov, Jeliazko R.; O'Meara, Matthew J.; DiMaio, Frank; Park, Hahnbeom; Shapovalov, Maxim V.; Renfrew, P. Douglas; Mulligan, Vikram Khipple; Kappel, Kalli; Labonte, Jason W.; Pacella, Michael S.; Bonneau, Richard; Bradley, Philip; Dunbrack, Roland L.; Das, Rhiju; Baker, David; Kuhlman, Brian; Kortemme, Tanja; Gray, Jeffrey J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. Journal of chemical theory and computation *13*, 3031-3048. 10.1021/acs.jctc.7b00125.

391. Robert, X.G., Patrice (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic acids research *42*, 320-324. 10.1093/nar/gku316.

392. Letunic, I.B., Peer (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research *49*, W293-W296. 10.1093/nar/gkab301.

393. Goddard, T.D.H., Conrad C.; Meng, Elaine C.; Pettersen, Eric F.; Couch, Gregory S.; Morris, John H.; Ferrin, Thomas E. (2017). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein science : a publication of the Protein Society *27*, 14-25. 10.1002/pro.3235.

394. Pettersen, E.F.G., Thomas D.; Huang, Conrad C.; Meng, Elaine C.; Couch, Gregory S.; Croll, Tristan I.; Morris, John H.; Ferrin, Thomas E. (2020). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein science : a publication of the Protein Society *30*, 70-82. 10.1002/pro.3943.

395. Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., et al. (2003). The Genome sequence of the SARS-associated coronavirus. Science *300*, 1399-1404. 10.1126/science.1085953.

396. Zhu, Y., Liu, M., Zhao, W., Zhang, J., Zhang, X., Wang, K., Gu, C., Wu, K., Li, Y., Zheng, C., et al. (2005). Isolation of virus from a SARS patient and genome-wide analysis of genetic mutations

related to pathogenesis and epidemiology from 47 SARS-CoV isolates. Virus Genes *30*, 93-102. 10.1007/s11262-004-4586-9.

397. Wolfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Muller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. Nature *581*, 465-469. 10.1038/s41586-020-2196-x.

398. Ng, D.L., Goldgof, G.M., Shy, B.R., Levine, A.G., Balcerek, J., Bapat, S.P., Prostko, J., Rodgers, M., Coller, K., Pearce, S., et al. (2020). SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood. Nature communications *11*, 4698. 10.1038/s41467-020-18468-8.

399. Hachim, A., Kavian, N., Cohen, C.A., Chin, A.W.H., Chu, D.K.W., Mok, C.K.P., Tsang, O.T.Y., Yeung, Y.C., Perera, R., Poon, L.L.M., et al. (2020). ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. Nat Immunol *21*, 1293-1301. 10.1038/s41590-020-0773-7.

400. Huang, A.T., Garcia-Carreras, B., Hitchings, M.D.T., Yang, B., Katzelnick, L.C., Rattigan, S.M., Borgert, B.A., Moreno, C.A., Solomon, B.D., Trimmer-Smith, L., et al. (2020). A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. Nature communications *11*, 4704. 10.1038/s41467-020-18450-4.

401. Poh, C.M., Carissimo, G., Wang, B., Amrun, S.N., Lee, C.Y., Chee, R.S., Fong, S.W., Yeo, N.K., Lee, W.H., Torres-Ruesta, A., et al. (2020). Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nature communications *11*, 2806. 10.1038/s41467-020-16638-2.

402. Liu, W., Liu, L., Kou, G., Zheng, Y., Ding, Y., Ni, W., Wang, Q., Tan, L., Wu, W., Tang, S., et al. (2020). Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. J Clin Microbiol *58*. 10.1128/JCM.00461-20.

403. Rikhtegaran Tehrani, Z., Saadat, S., Saleh, E., Ouyang, X., Constantine, N., DeVico, A.L., Harris, A.D., Lewis, G.K., Kottilil, S., and Sajadi, M.M. (2020). Performance of nucleocapsid and spike-based SARS-CoV-2 serologic assays. PLoS One *15*, e0237828. 10.1371/journal.pone.0237828.

404. Jiang, H.W., Li, Y., Zhang, H.N., Wang, W., Yang, X., Qi, H., Li, H., Men, D., Zhou, J., and Tao, S.C. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. Nature communications *11*, 3581. 10.1038/s41467-020-17488-8.

405. Di, B., Hao, W., Gao, Y., Wang, M., Wang, Y.D., Qiu, L.W., Wen, K., Zhou, D.H., Wu, X.W., Lu, E.J., et al. (2005). Monoclonal antibody-based antigen capture enzyme-linked immunosorbent assay reveals high sensitivity of the nucleocapsid protein in acute-phase sera of severe acute respiratory syndrome patients. Clin Diagn Lab Immunol *12*, 135-140. 10.1128/CDLI.12.1.135-140.2005.

406. Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., et al. (2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem Biophys Res Commun *527*, 618-623. 10.1016/j.bbrc.2020.04.136.

407. Chang, C.K., Sue, S.C., Yu, T.H., Hsieh, C.M., Tsai, C.K., Chiang, Y.C., Lee, S.J., Hsiao, H.H., Wu, W.J., Chang, W.L., et al. (2006). Modular organization of SARS coronavirus nucleocapsid protein. J Biomed Sci *13*, 59-72. 10.1007/s11373-005-9035-9.

408. Peng, Y., Du, N., Lei, Y., Dorje, S., Qi, J., Luo, T., Gao, G.F., and Song, H. (2020). Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. The EMBO journal *39*, e105938. 10.15252/embj.2020105938.

409. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., et al. (2020). Molecular Architecture of the SARS-CoV-2 Virus. Cell *183*, 730-738 e713. 10.1016/j.cell.2020.09.018.

410. Gui, M., Liu, X., Guo, D., Zhang, Z., Yin, C.C., Chen, Y., and Xiang, Y. (2017). Electron microscopy studies of the coronavirus ribonucleoprotein complex. Protein Cell *8*, 219-224. 10.1007/s13238-016-0352-8.

411. Li, T., Wang, L., Wang, H., Li, X., Zhang, S., Xu, Y., and Wei, W. (2020). Serum SARS-COV-2 Nucleocapsid Protein: A Sensitivity and Specificity Early Diagnostic Marker for SARS-COV-2 Infection. Front Cell Infect Microbiol *10*, 470. 10.3389/fcimb.2020.00470.

412. Che, X.Y., Hao, W., Wang, Y., Di, B., Yin, K., Xu, Y.C., Feng, C.S., Wan, Z.Y., Cheng, V.C., and Yuen, K.Y. (2004). Nucleocapsid protein as early diagnostic marker for SARS. Emerg Infect Dis *10*, 1947-1949. 10.3201/eid1011.040516.

413. Diao, B., Wen, K., Zhang, J., Chen, J., Han, C., Chen, Y., Wang, S., Deng, G., Zhou, H., and Wu, Y. (2020). Accuracy of a nucleocapsid protein antigen rapid test in the diagnosis of SARS-CoV-2 infection. Clin Microbiol Infect. 10.1016/j.cmi.2020.09.057.

414. Grant, B.D., Anderson, C.E., Williford, J.R., Alonzo, L.F., Glukhova, V.A., Boyle, D.S., Weigl, B.H., and Nichols, K.P. (2020). SARS-CoV-2 Coronavirus Nucleocapsid Antigen-Detecting Half-Strip Lateral Flow Assay Toward the Development of Point of Care Tests Using Commercially Available Reagents. Anal Chem *92*, 11305-11309. 10.1021/acs.analchem.0c01975.

415. Murphy, K., Travers, P., Walport, M., Janeway, C. (2008). Janeway's immunobiology (Garland Science).

416. Knutson, D.W., van Es, L.A., Kayser, B.S., and Glassock, R.J. (1979). Soluble oligovalent antigen--antibody complexes. II. The effect of various selective forces upon relative stability of isolated complexes. Immunology *37*, 495-503.

417. van Es, L.A., Knutson, D.W., Kayser, B.S., and Glassock, R.J. (1979). Soluble oligovalent antigen-antibody complexes. I. The effect of antigen valence and combining ratio on the composition of fluorescein-carrier anti-fluorescein complexes. Immunology *37*, 485-493.

418. Qiu, L.W., Tang, H.W., Wang, Y.D., Liao, J.E., Hao, W., Wen, K., He, X.M., and Che, X.Y. (2005). [Development and application of triple antibodies-based sandwich ELISA for detecting nucleocapsid protein of SARS-associated coronavirus]. Zhonghua Liu Xing Bing Xue Za Zhi *26*, 277-281.

419. Bongini, L., Fanelli, D., Piazza, F., De Los Rios, P., Sandin, S., and Skoglund, U. (2005). Dynamics of antibodies from cryo-electron tomography. Biophys Chem *115*, 235-240. 10.1016/j.bpc.2004.12.037.

420. Sandin, S., Ofverstedt, L.G., Wikstrom, A.C., Wrange, O., and Skoglund, U. (2004). Structure and flexibility of individual immunoglobulin G molecules in solution. Structure *12*, 409-415. 10.1016/j.str.2004.02.011.

421. Jay, J.W., Bray, B., Qi, Y., Igbinigie, E., Wu, H., Li, J., and Ren, G. (2018). IgG Antibody 3D Structures and Dynamics. Antibodies (Basel) *7*. 10.3390/antib7020018.

422.   Lei, D., Liu, J., Liu, H., Cleveland, T.E.t., Marino, J.P., Lei, M., and Ren, G. (2019). Single-Molecule 3D Images of "Hole-Hole" IgG1 Homodimers by Individual-Particle Electron Tomography. Scientific reports *9*, 8864. 10.1038/s41598-019-44978-7.

423.   Zhang, X., Zhang, L., Tong, H., Peng, B., Rames, M.J., Zhang, S., and Ren, G. (2015). 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. Sci Rep *5*, 9803. 10.1038/srep09803.

424.   Tian, X., Vestergaard, B., Thorolfsson, M., Yang, Z., Rasmussen, H.B., and Langkilde, A.E. (2015). In-depth analysis of subclass-specific conformational preferences of IgG antibodies. IUCrJ *2*, 9-18. 10.1107/S205225251402209X.

425.   Wright, D.W., Elliston, E.L.K., Hui, G.K., and Perkins, S.J. (2019). Atomistic Modeling of Scattering Curves for Human IgG1/4 Reveals New Structure-Function Insights. Biophys J *117*, 2101-2119. 10.1016/j.bpj.2019.10.024.

426.   Yanaka, S., Yogo, R., and Kato, K. (2020). Biophysical characterization of dynamic structures of immunoglobulin G. Biophys Rev *12*, 637-645. 10.1007/s12551-020-00698-1.

427.   Remesh, S.G., Armstrong, A.A., Mahan, A.D., Luo, J., and Hammel, M. (2018). Conformational Plasticity of the Immunoglobulin Fc Domain in Solution. Structure *26*, 1007-1014 e1002. 10.1016/j.str.2018.03.017.

428.   Liu, X., Zhao, Y., Shi, H., Zhang, Y., Yin, X., Liu, M., Zhang, H., He, Y., Lu, B., Jin, T., and Li, F. (2019). Human immunoglobulin G hinge regulates agonistic anti-CD40 immunostimulatory and antitumour activities through biophysical flexibility. Nature communications *10*, 4206. 10.1038/s41467-019-12097-6.

429.   White, A.L., Chan, H.T., French, R.R., Willoughby, J., Mockridge, C.I., Roghanian, A., Penfold, C.A., Booth, S.G., Dodhy, A., Polak, M.E., et al. (2015). Conformation of the human immunoglobulin G2 hinge imparts superagonistic properties to immunostimulatory anticancer antibodies. Cancer Cell *27*, 138-148. 10.1016/j.ccell.2014.11.001.

430.   Mayes, P.A., Hance, K.W., and Hoos, A. (2018). The promise and challenges of immune agonist antibody development in cancer. Nat. Rev. Drug Discov. *17*, 509-527. 10.1038/nrd.2018.75.

431.   Pelikan, M., Hura, G.L., and Hammel, M. (2009). Structure and flexibility within proteins as identified through small angle X-ray scattering. Gen. Physiol. Biophys. *28*, 174-189.

432.   Zhao, J., Nussinov, R., and Ma, B. (2019). Antigen binding allosterically promotes Fc receptor recognition. MAbs *11*, 58-74. 10.1080/19420862.2018.1522178.

433.   Patterson, E.I., Prince, T., Anderson, E.R., Casas-Sanchez, A., Smith, S.L., Cansado-Utrilla, C., Solomon, T., Griffiths, M.J., Acosta-Serrano, A., Turtle, L., and Hughes, G.L. (2020). Methods of Inactivation of SARS-CoV-2 for Downstream Biological Assays. J Infect Dis *222*, 1462-1467. 10.1093/infdis/jiaa507.

434.   Alpdagtas, S., Ilhan, E., Uysal, E., Sengor, M., Ustundag, C.B., and Gunduz, O. (2020). Evaluation of current diagnostic methods for COVID-19. APL Bioeng *4*, 041506. 10.1063/5.0021554.

435.   FDA (2020). Coronavirus (COVID-19) Update: FDA Authorizes Antigen Test as First Over-the-Counter Fully At-Home Diagnostic Test for COVID-19.

436. Wang, D., He, S., Wang, X., Yan, Y., Liu, J., Wu, S., Liu, S., Lei, Y., Chen, M., Li, L., et al. (2020). Rapid lateral flow immunoassay for the fluorescence detection of SARS-CoV-2 RNA. Nat Biomed Eng *4*, 1150-1158. 10.1038/s41551-020-00655-z.

437. Ahn, D.G., Jeon, I.J., Kim, J.D., Song, M.S., Han, S.R., Lee, S.W., Jung, H., and Oh, J.W. (2009). RNA aptamer-based sensitive detection of SARS coronavirus nucleocapsid protein. Analyst *134*, 1896-1901. 10.1039/b906788d.

438. Loeffelholz, M.J., and Tang, Y.W. (2020). Laboratory diagnosis of emerging human coronavirus infections - the state of the art. Emerg Microbes Infect *9*, 747-756. 10.1080/22221751.2020.1745095.

439. Hesari, M., and Ding, Z. (2020). A Perspective on Application of Carbon Quantum Dots in Luminescence Immunoassays. Front Chem *8*, 580033. 10.3389/fchem.2020.580033.

440. Teengam, P., Siangproh, W., Tuantranont, A., Vilaivan, T., Chailapakul, O., and Henry, C.S. (2017). Multiplex Paper-Based Colorimetric DNA Sensor Using Pyrrolidinyl Peptide Nucleic Acid-Induced AgNPs Aggregation for Detecting MERS-CoV, MTB, and HPV Oligonucleotides. Anal Chem *89*, 5428-5435. 10.1021/acs.analchem.7b00255.

441. Fabiani, L., Saroglia, M., Galata, G., De Santis, R., Fillo, S., Luca, V., Faggioni, G., D'Amore, N., Regalbuto, E., Salvatori, P., et al. (2021). Magnetic beads combined with carbon black-based screen-printed electrodes for COVID-19: A reliable and miniaturized electrochemical immunosensor for SARS-CoV-2 detection in saliva. Biosens Bioelectron *171*, 112686. 10.1016/j.bios.2020.112686.

442. Yagi, H., Yanaka, S., and Kato, K. (2018). Structure and Dynamics of Immunoglobulin G Glycoproteins. Adv. Exp. Med. Biol. *1104*, 219-235. 10.1007/978-981-13-2158-0_11.

443. Lin, A.V. (2015). Direct ELISA. Methods Mol Biol *1318*, 61-67. 10.1007/978-1-4939-2742-5_6.

444. Tang, D., Sauceda, J.C., Lin, Z., Ott, S., Basova, E., Goryacheva, I., Biselli, S., Lin, J., Niessner, R., and Knopp, D. (2009). Magnetic nanogold microspheres-based lateral-flow immunodipstick for rapid detection of aflatoxin B2 in food. Biosens Bioelectron *25*, 514-518. 10.1016/j.bios.2009.07.030.

445. Zhang, S., Garcia-D'Angeli, A., Brennan, J.P., and Huo, Q. (2014). Predicting detection limits of enzyme-linked immunosorbent assay (ELISA) and bioanalytical techniques in general. Analyst *139*, 439-445. 10.1039/c3an01835k.

446. Eschenfeldt, W.H., Makowska-Grzyska, M., Stols, L., Donnelly, M.I., Jedrzejczak, R., and Joachimiak, A. (2013). New LIC vectors for production of proteins from genes containing rare codons. J Struct Funct Genomics *14*, 135-144. 10.1007/s10969-013-9163-9.

447. Saphire, E.O., Parren, P.W., Pantophlet, R., Zwick, M.B., Morris, G.M., Rudd, P.M., Dwek, R.A., Stanfield, R.L., Burton, D.R., and Wilson, I.A. (2001). Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. Science *293*, 1155-1159. 10.1126/science.1061692 293/5532/1155 [pii].

448. Hammel, M., Kriechbaum, M., Gries, A., Kostner, G.M., Laggner, P., and Prassl, R. (2002). Solution Structure of Human and Bovine β2-Glycoprotein I Revealed by Small-angle X-ray Scattering. J. Mol. Biol. *321*, 85-97. 10.1016/s0022-2836(02)00621-6.

449. Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr *42*, 342-346. Doi 10.1107/S0021889809000338.

450.   Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. J. Appl. Cryst. *36*, 860-864.

451.   Wriggers, W., Milligan, R.A., and McCammon, J.A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. Journal of structural biology *125*, 185-195.

452.   Booth, D.S., Avila-Sakar, A., and Cheng, Y. (2011). Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition. J Vis Exp. 10.3791/3227.

453.   Scarff, C.A., Fuller, M.J.G., Thompson, R.F., and Iadanza, M.G. (2018). Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. J Vis Exp. 10.3791/57199.

454.   Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519-530. 10.1016/j.jsb.2012.09.006.

455.   Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*. 10.7554/eLife.42166.

456.   Yin, W.C., Mao, C.Y., Luan, X.D., Shen, D.D., Shen, Q.Y., Su, H.X., Wang, X.X., Zhou, F.L., Zhao, W.F., Gao, M.Q., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. Science *368*, 1499-+. 10.1126/science.abc1560.

457.   Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell *181*, 914-+. 10.1016/j.cell.2020.04.011.

458.   Ahn, D.G., Choi, J.K., Taylor, D.R., and Oh, J.W. (2012). Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. Arch Virol *157*, 2095-2104. 10.1007/s00705-012-1404-x.

459.   Gao, Y., Yan, L.M., Huang, Y.C., Liu, F.J., Zhao, Y., Cao, L., Wang, T., Sun, Q.Q., Ming, Z.H., Zhang, L.Q., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science *368*, 779-+. 10.1126/science.abb7498.

460.   Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B., and Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. P Natl Acad Sci USA *111*, E3900-E3909. 10.1073/pnas.1323705111.

461.   Snijder, E.J., Decroly, E., and Ziebuhr, J. (2016). The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. Adv Virus Res *96*, 59-126. 10.1016/bs.aivir.2016.08.008.

462.   Tvarogova, J., Madhugiri, R., Bylapudi, G., Ferguson, L.J., Karl, N., and Ziebuhr, J. (2019). Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3 '-Terminal Adenylyltransferase Activity. J Virol *93*. ARTN e00291-19 10.1128/JVI.00291-19.

463.   Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., et al. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. Cell. 10.1016/j.cell.2020.10.004.

464. Zhai, Y.J., Sun, F., Li, X.M., Pang, H., Xu, X.L., Bartlam, M., and Rao, Z.H. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol *12*, 980-986. 10.1038/nsmb999.

465. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature *584*, 154-+. 10.1038/s41586-020-2368-8.

466. Peng, Q., Peng, R.C., Yuan, B., Zhao, J.R., Wang, M., Wang, X.X., Wang, Q., Sun, Y., Fan, Z., Qi, J.X., et al. (2020). Structural and Biochemical Characterization of the nsp12-nsp7-nsp8 Core Polymerase Complex from SARS-CoV-2. Cell Rep *31*. UNSP 107774 10.1016/j.celrep.2020.107774.

467. Yan, L.M., Zhang, Y., Ge, J., Zheng, L.T., Gao, Y., Wang, T., Jia, Z.H., Wang, H.F., Huang, Y.C., Li, M.Y., et al. (2020). Architecture of a SARS-CoV-2 mini replication and transcription complex. Nat. Commun. *11*, 6, 5874. 10.1038/s41467-020-19770-1.

468. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T., et al. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell *182*, 1560-+. 10.1016/j.cell.2020.07.033.

469. Liming Yan, J.G., Litao Zheng, Ying Zhang, Yan Gao, Tao Wang, Yucen Huang, Yunxiang Yang, Shan Gao,, and Mingyu Li, Z.L., Haofeng Wang, Yingjian Li, Yu Chen, Luke W. Guddat, Quan Wang, Zihe Rao, Zhiyong Lou (2021). Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. Cell *184*, 184–193.

470. Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020). Structural analysis of the putative SARS-CoV-2 primase complex. J Struct Biol *211*. ARTN 107548 UNSP 107548 10.1016/j.jsb.2020.107548.

471. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature. 10.1038/s41586-020-2368-8.

472. Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J Appl Crystallogr *42*, 342-346. 10.1107/S0021889809000338.

473. Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G., and Hilgenfeld, R. (2012). Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. J Virol *86*, 4444-4454. 10.1128/JVI.06635-11.

474. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041. 10.1073/pnas.181342398 181342398 [pii].

475. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., and Baker, N.A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic acids research *32*, W665-667. 10.1093/nar/gkh381.

476. Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., and Rao, Z. (2005). Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol *12*, 980-986. 10.1038/nsmb999.

477. Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. Science *336*, 1030-1033. 10.1126/science.1218231.

478. Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*, W615-619. 10.1093/nar/gkh39832/suppl_2/W615 [pii].

479. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2011). Macromolecular docking restrained by a small angle X-ray scattering profile. Journal of structural biology *173*, 461-471. S1047-8477(10)00292-3 [pii] 10.1016/j.jsb.2010.09.023.

480. Aslanidis, C., and de Jong, P.J. (1990). Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res *18*, 6069-6074. 10.1093/nar/18.20.6069.

481. Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR *20*, 71-75. 10.1023/a:1011254402785.

482. Sherlin, L.D., Bullock, T.L., Nissan, T.A., Perona, J.J., Lariviere, F.J., Uhlenbeck, O.C., and Scaringe, S.A. (2001). Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. RNA *7*, 1671-1678.

483. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. Acta Crystallogr D *62*, 859-866.

484. French, S., and Wilson, K. (1978). Treatment of Negative Intensity Observations. Acta Crystallogr., Sect. A: Found. Crystallogr. *34*, 517-525. 10.1107/S0567739478001114.

485. Padilla, J.E., and Yeates, T.O. (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. Acta Crystallogr., Sect D: Biol. Crystallogr. *59*, 1124-1130.

486. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr., Sect D: Biol. Crystallogr. *67*, 235-242. S0907444910045749 [pii] 10.1107/S0907444910045749.

487. Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr., Sect D: Biol. Crystallogr. *66*, 22-25. 10.1107/S0907444909042589.

488. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr., Sect D: Biol. Crystallogr. *53*, 240-255. 10.1107/S0907444996012255 S0907444996012255 [pii].

489. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., Sect D: Biol. Crystallogr. *66*, 213-221. S0907444909052925 [pii] 10.1107/S0907444909052925.

490. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography *26*, 283-291. doi:10.1107/S0021889892009944.

491. Guinier, A., and Fournet, F. (1955). Small angle scattering of X-rays (Wiley Interscience).

492. Heller, W.T., Urban, V.S., Lynn, G.W., Weiss, K.L., O'Neill, H.M., Pingali, S.V., Qian, S., Littrell, K.C., Melnichenko, Y.B., Buchanan, M.V., et al. (2014). The Bio-SANS instrument at

the High Flux Isotope Reactor of Oak Ridge National Laboratory. Journal of Applied Crystallography *47*, 1238-1246. doi:10.1107/S1600576714011285.

493. Heller, W.T. (2010). Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. Acta Crystallographica Section D *66*, 1213-1217. 10.1107/S0907444910017658.

494. Hopkins, J.B., Gillilan, R.E., and Skou, S. (2017). BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. Journal of applied crystallography *50*, 1545-1553. 10.1107/S1600576717011438.

495. Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., and Svergun, D.I. (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. Journal of applied crystallography *50*, 1212-1225. 10.1107/S1600576717007786.

496. Piiadov, V., Ares de Araujo, E., Oliveira Neto, M., Craievich, A.F., and Polikarpov, I. (2019). SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. Protein Sci *28*, 454-463. 10.1002/pro.3528.

497. Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *76*, 2879-2886.

498. Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr *45*, 342-350. 10.1107/S0021889812007662.

499. Petoukhov, M.V., and Svergun, D.I. (2006). Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. Eur Biophys J *35*, 567-576. 10.1007/s00249-006-0063-9.

500. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr *36*, 1277-1282. 10.1107/s0021889803012779.

501. Svergun, D.I., Richard, S., Koch, M.H., Sayers, Z., Kuprin, S., and Zaccai, G. (1998). Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc Natl Acad Sci U S A *95*, 2267-2272. 10.1073/pnas.95.5.2267.