INVESTIGATING THE ROLE OF ALLOSTERY THROUGH CHANGES IN PROTEIN STABILITY AND DYNAMICS

by

Angela Jean Patterson

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2021

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor and mentor Dr. Brian Bothner for giving me the guidance needed and the opportunity to be successful in my graduate career.

I would also like to thank my graduate committee: Dr. Brian Bothner, Dr. Blake Wiedenheft, Dr. Martin Lawrence, Dr. Valérie Copié, and Dr. Brian Tripet. They have given me the opportunity to begin to think deeply about proteins and how they function. Whether through in class lectures or collaboration, I have learned a great deal about the fundamentals of protein structure and function and the relationship between the two from my committee members.

I would like to thank the Bothner lab members as a whole. The work environment provided by my peers has been one of collaboration and inspiration for me, and each lab member participated in thoughtful discussions whether it be technique based or biological system based throughout my graduate studies. In particular, I would like to thank Monika Tokmina-Lukaszewska for her guidance and patience in teaching me everything I know about mass spectrometry and all of the technical aspects of data analysis. I would like to thank Ravi Kant, Luke Berry, and Jenna Mattice for their thoughtful discussions on HDX-MS as well as their physical help in running these large experiments. I would like to thank all of the undergraduates who have worked with me throughout my time in graduate school. Each of you has worked hard to support the projects presented in this thesis.

Lastly, I would like to thank my family and friends. Without your support, I would have never been successful in graduate school. Thank you.

TABLE OF CONTENTS

1. INTRODUCTION 1

Protein Dynamics 3
Ligand Binding Alters the Conformational Landscape 5
Allostery 6
Introduction to HDX-MS 10
Applying HDX-MS to Track Allosteric Networks 13
References 15

2. HYDROGEN DEUTERIUM EXCHANGE REVEALS A DYNAMIC DNA-BINDING MAP OF REPLICATION PROTEIN A 18

Contribution of Authors and Co-Authors 18
Manuscript Information 20
Abstract 21
Introduction 21
Material and Methods 27
HDX-MS 27
Data Analysis 28
Results 29
HDX-MS Generates a Conformational Landscape for DNA-Induced Changes in Human RPA 29
The FAB Half of RPA Shows Highly Dynamic ssDNA Dependent Changes in HDX 34
Tri-C Shows Less Dynamic ssDNA Dependent Changes in HDX 38
Discussion 47
References 52

3. DYNAMICS OF HEPATITIS B VIRUS CAPSID PROTEIN DIMER REGULATES ASSEMBLY THROUGH AN ALLOSTERIC NETWORK 58

Contribution of Authors and Co-Authors 58
Manuscript Information 59
Abstract 60
Introduction 61
Methods 65
Purification of HBV and WHV Dimers 65

TABLE OF CONTENTS CONTINUED

Differential Scanning Fluorimetry (DSF) of HBV and WHV Dimers 66
Native Mass Spectrometry (NMS) of HBV and WHV Dimers 66
Hydrogen Deuterium Exchange Coupled to Mass Spectrometry (HDX-MS) of HBV and WHV Dimers 67
Results 68
Differential Scanning Fluorimetry (DSF) of HBV and WHV Dimers 68
Native Mass Spectrometry Probes Intra-Dimer Interface Stability 70
HDX-MS of wCp149 and hCp149 Dimers 71
HDX-MS Analysis of Mutant Cp149 73
Discussion 75
Stability of the HBV Intra-Dimer Interface 76
Inter-Dimer Contact Region of WHV and HBV 78
Comparison of Dynamics Between the hCp149 Free Dimer and hCp149 Capsid 79
Conclusions 80
References 81

4. CONFORMATIONAL DYNAMICS OF DNA BINDING AND CAS3 RECRUITMENT BY THE CRISPR RNA-GUIDE CASCADE COMPLEX 86

Contribution of Authors and Co-Authors 86
Manuscript Information 88
Abstract 89
Introduction 89
Results 93
DNA Binding 96
Path of the Displaced Strand 98
Recruitment of the Nuclease-Helicase Cas3 102
Discussion 106
Materials and Methods 110
Hydrogen-Deuterium Exchange Coupled with Mass Spectrometry (HDX-MS) 110
Cascade Expression and Purification 113
Cas3 Expression and Purification 114
Oligoduplex Preparation and DNA Labeling 114

TABLE OF CONTENTS CONTINUED

Electrophoretic Mobility Shift Assay (EMSA)....115
Plasmid Curing Assays....115
Cascade/Cas3 Mediated DNA Degradation....116
References....117

5. THERMODYNAMICS OF CRISPR-ANTI-CRISPR INTERACTIONS PROVIDES MECHANISTIC INSIGHT INTO INHIBITION....122

Contribution of Authors and Co-Authors....122
Manuscript Information....124
Abstract....125
Introduction....126
Methods....130
Purification of the Csy Complex....130
ssDNA Bound to Csy....131
Purification of AcrIF9 Bound Csy....131
Purification of AcrIF2 Bound Csy....131
Intact HDX-MS....132
Peptide Level HDX-MS....133
Differential Scanning Fluorimetry....134
Results....135
Intact Protein HDX-MS Shows Global Differences in Protein Dynamics Upon Acr Binding....135
Peptide Level HDX-MS Shows Differences in Protein Dynamics Upon Acr Binding....138
ssDNA Binding Shows a Decrease in HDX Throughout the Csy Complex....140
Differences in Csy Thermal Stability Upon Acr Binding....141
Discussion....144
AcrIF9 Binding is Entropically Stabilized....145
AcrIF2 Binding is Enthalpically Stabilized....146
Binding of Acrs Alters the Conformational Landscape Implying a Role for Allostery....146
AcrIF9 Interactions with DNA....149
Conclusions....151
References....152

TABLE OF CONTENTS CONTINUED

6. SUMMARY 156

Allostery in RPA 157
Allostery in HBV 160
Allostery in Cascade 163
Allostery in Csy 165
Is Allostery Present in Every Protein Complex? 168
References 171

CUMULATIVE REFERENCES CITED 173

LIST OF TABLES

Table Page

1.1 Timescales and Distances Associated with Protein Movement.....4

LIST OF FIGURES

Figure Page

1.1 A 2D Cartoon Representation of a Protein Folding Funnel....2

1.2 Time Continuum of Protein Motions....5

1.3 Allowed States for Each of the Proposed Allosteric Models....7

1.4 A Typical HDX-MS Workflow....12

2.1 Sequence Coverage of the DNA Binding and Protein Interaction Domains of RPA....23

2.2 Global Deuterium Uptake Profiles and DNA Induced Changes in RPA....30

2.3 Time-Resolved HDX-MS Analysis of DNA-Induced Conformational Changes in the FAB Half of RPA....35

2.4 Time-Resolved HDX-MS Analysis of DNA-Induced Conformational Changes in DBD-C....40

2.5 Time-Resolved HDX-MS Analysis of DNA-Induced Conformational Changes in RPA32 and RPA14....44

2.6 RPA Binds to ssDNA Using Dynamics and Less Dynamic Halves....49

3.1 Structure of Human and Woodchuck Hepatitis B Viruses....62

3.2 Differential Scanning Fluorimetry of HBV and WHV Dimers....69

3.3 Native Mass Spectrometry of hCp149 and wCp149....71

3.4 Deuterium Uptake in the Intra-Dimer Interface....73

3.5 Mutation at Residue 132 Alters Exchange Across Cp149....75

4.1 Binding of dsDNA Increases the Stability of Cascade....90

LIST OF FIGURES CONTINUED

Figure Page

4.2 Peptide-Resolution Mapping of Deuterium Uptake Reveals Areas of Cascade that Become More or Less Protected Upon dsDNA Binding 95

4.3 Using HDX to Identify Residues Involved in Binding the Displaced DNA Strand....99

4.4 Predicting the Cas3 Interaction Site Using HDX 104

5.1 Cartoon Depictions of the Csy Complex Either Bound to AcrIF9 or AcrIF2 127

5.2 Protein Level HDX Shows Differences in Dynamics of the Csy Complex When Different anti-CRISPRs Are Bound 137

5.3 Peptide Level HDX Shows Differences in Dynamics of the Csy Complex When Different anti-CRISPRs Are Bound 139

5.4 Peptide Level and Protein Level HDX Shows Decreased HDX Upon the Binding of ssDNA 141

5.5 Differential Scanning Fluorimetry Data for the Unbound, ssDNA AcrIF9, and AcrIF2 Bound Csy Complex 142

5.6 The Conformational Energy Landscape is Remodeled Upon Ligand Binding 148

5.7 AcrIF9 Changes the Conformation and Deuterium Exchange in the Cas8f N-Terminal Hook Region 151

6.1 Schematic Representation of the Allosteric Network and Energetic Effects Observed Upon ssDNA Binding 158

6.2 Representation of the Effects of the Y132A Mutation on the Protein Dynamics of the Cp149 Dimer as well as the Energy Landscape 161

6.3 Schematic Representation of the Allosteric Network Within Cse1 of the Cascade Complex and its Effects on the Energy Landscape 164

LIST OF FIGURES CONTINUED

Figure Page

6.4 Allosteric Effects and the Changes in the Energy Landscape Caused by Acr Binding.....167

6.5 Summary of Findings Covered in this Thesis.....169

ABSTRACT

Allostery is the presence of a communication network that links functional sites of a protein that are distal from one another. The existence of an allosteric network can be observed through conformational change or a change in protein dynamics. These networks can be used to provide insight into the mechanistic function of proteins or protein complexes. In this thesis, four protein complexes were studied (RPA, HBV, Cascade, and Csy) and allosteric networks within the complexes were observed by monitoring the changes in protein dynamics upon an energy perturbation. To measure the changes in protein dynamics, hydrogen deuterium exchange mass spectrometry was used. This technique allows for the determination of how often the hydrogen bonding within a protein structure is broken. By tracking the longevity of the hydrogen bonding network that comprises the studied protein's structure, the dynamics of the protein can be studied. In this work, each of the proteins had changes in protein dynamics that were distal from the site of the energy perturbation that had functional impacts on each of the protein complexes. The combined presence of the distal changes in dynamics with an effect on protein function fits the definition of allostery. If allostery is present in these four diverse systems, is it possible that allostery is present in all proteins?

CHAPTER ONE

INTRODUCTION

According to Levinthal's Paradox, the amount of time it would take for a protein to fold if it randomly sampled every possible configuration would be too long to be biologically relevant *(1)*. There are two models to explain how protein folding occurs. In one model there is a single pathway that the protein takes to the final structure and there are set intermediates in this folding pathway *(2)*. However, this model does not explain the presence of differently folded intermediate species. This led to the energy landscape theory of protein folding. The energy landscape theory was adopted to explain how proteins fold in microseconds to milliseconds, as well as the presence of unpredictable intermediates *(3,4)*. As proteins fold they travel down a rugged energy landscape that narrows as the protein approaches its native fold; this is termed a protein folding funnel *(5,6)* (Figure 1.1). At the top of the funnel, the sample space is large so denatured proteins can adopt many conformations, but as the protein begins to fold, it travels down the energy landscape (Z-axis) and the available number of conformations lessens (X and Y axes) *(6)*.

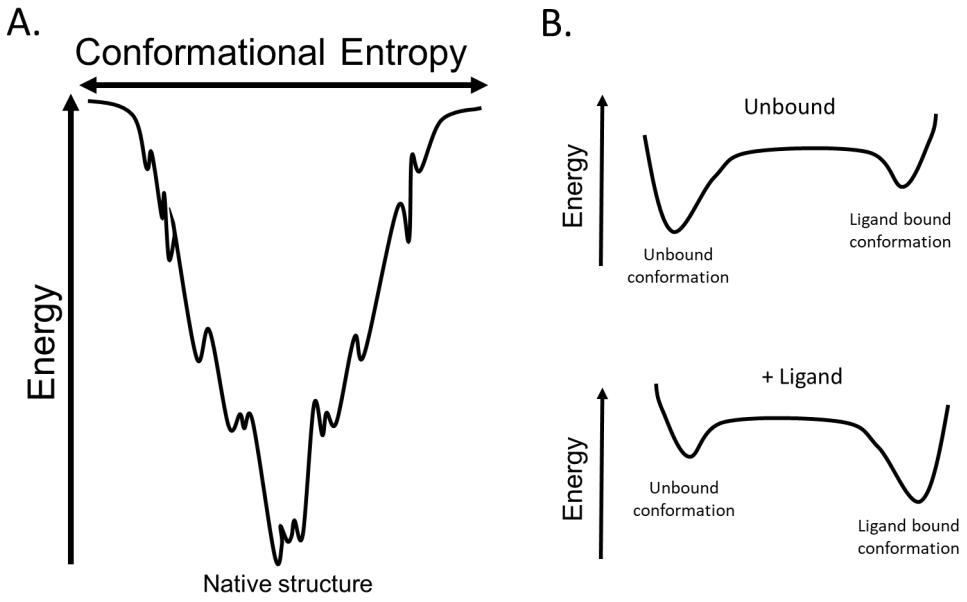

Figure 1.1. A 2D cartoon representation of a protein folding funnel. (A) The wider the x and y axes, the more conformational entropy is present. The funnel is wider toward higher energy conformations. As the protein folds, it loses entropy but this is compensated for by a release of energy (favorable enthalpy). As the protein travels down the funnel, the conformational landscape becomes more narrow allowing for fewer conversions between conformations of similar energy. The conformational space continues to narrow until the base of the funnel is reached. The base of the funnel can be a smooth valley or a rugged valley allowing for fewer or more native conformations in the ensemble. Figure adapted from Wolynes *et al.* (1995). (B) The base of the protein folding funnel showing two possible conformations in the native ensemble. One is the most prevalent conformation in the absence of ligand and the other is the most populated conformation in the presence of ligand. The higher in energy the conformational well lies, the less the conformation is represented in the conformational ensemble. Higher energy wells contain fewer representative proteins than lower energy wells. In the unbound condition, both conformations can be sampled, though the unbound conformation will be sampled with higher frequency (this conformation will be more populated) than the ligand bound conformation. However, ligand binding remodels the energy landscape so that the ligand bound conformation lies at a lower energy and is therefore more populated, but both conformations are represented in each scenario.

The bottom of the protein folding funnel may not be a smooth valley and instead the native form of a protein may exist as a conformational ensemble. Each conformation may be represented as a shallow valley that exists at the same energy and is separated

from other valleys by a low activation energy barrier for interconversion *(4)*. Higher energy valleys may also be present, and these conformations may still be sampled but with less frequency. The population distribution of the conformational ensemble is dependent on the energy of the conformation with the lowest energy conformations making up the majority of the distribution of the ensemble.

Protein Dynamics

Conversion between conformational states within the energy landscape leads to protein dynamics. There is movement associated with the protein converting between accessible conformations. More dynamic proteins (such as intrinsically disordered proteins) sample many similar in energy conformations and more stable proteins (such as structural proteins) sample fewer conformations.

Protein dynamics can be described as a continuum, with intrinsically disordered proteins (IDPs) or regions (IDRs) possessing the most dynamic nature at one end and strongly folded domains are at the other *(7)*. IDPs and IDRs possess a wide conformational space leading to the sampling of many conformations, thus making them highly dynamic. Local unfolding events could be considered the next most dynamic stage. These events also lead to a variety of conformations that can be adopted within the unfolded region causing these regions to be highly dynamic. The next range of dynamics could be considered backbone dynamics, this leads to larger scale conformational changes. Following this, would be side chain dynamics. These would alter the conformation in local areas. Finally, in the continuum are rigid body motions. This is a

rearrangement of domains with respect to one another, where hinge regions would show protein dynamics, but the structured domains remain static.

These different motions can span a wide range of timescales (10^{-15} seconds to days) *(3)*. Figure 1.2 gives ranges of protein motion as a time continuum. The distance of these motions also has a wide range: μm (bond vibration) to nm (local denaturation) *(8)*. Larger motions have more time associated with them (local denaturation, 0.5-1 nm, milliseconds-seconds timescale) and more local motions occur on a faster timescale (bond vibration, 0.001-0.01nm, 10^{-14} seconds) *(8)*. Table 1.1 provides insight into the timescales of protein motions as well as the distance traveled during the motion. The conformational changes that were probed in this thesis occurred on the seconds to minutes timescale.

PROTEIN MOVEMENT	DISTANCE OF MOVEMENT (NANOMETERS)	TIMESCALE OF MOVEMENT (SECONDS)
Bond vibration	0.001-0.01	10^{-14}-10^{-13}
Side chain rotation	0.5-1.0	10^{-11}-10^{-10}
Motions of globular domains	0.1-0.5	10^{-10}-10^{-7}
Rigid body motions	1.0-5.0	10^{-9}-10^{-6}
Helix-coil transition	>3.0	10^{-7}-10^{-5}
Beta sheet formation	>2.0	10^{-6}-10^{-3}
Local denaturation	0.5-1.0	10^{-5}-10^{1}

Table 1.1 Timescales and distances associated with protein movement. Table adapted from reference (8).

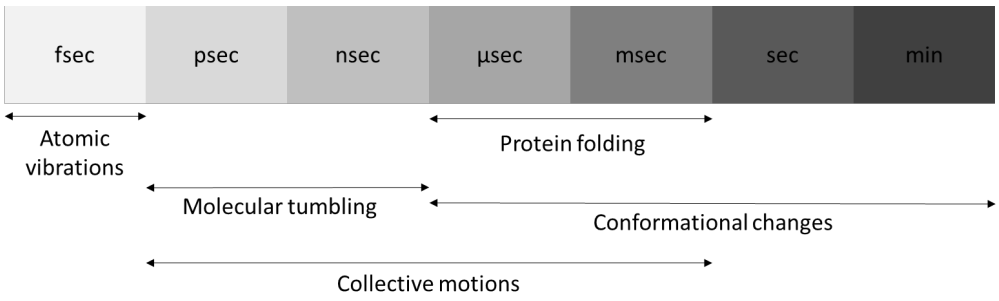

Figure 1.2 Time continuum of protein motions. Conformational changes probed in the following chapters of this thesis occur on the seconds to minutes timescale. Figure adapted from Ben-Nissan and Sharon (2011).

Ligand Binding Alters the Conformational Landscape

The rugged base of the protein folding funnel also has implications in protein binding. Originally it was thought that the active site of an enzyme was the exact shape of the substrate that binds (key and lock model), however, this did not explain differing affinities for similarly structured substrates *(9)*. This led to the development of the induced fit model of ligand binding. In the induced fit model, the interaction between the ligand and the protein remodels the intramolecular interactions of the protein thus causing conformational change to occur as ligand binds *(9)*. This model explains the presence of a ligand bound conformation which differs from the unbound conformation. However, if proteins exist as an ensemble of conformations, it would be logical that both conformations are sampled to some extent by the protein. To understand how ligand binding effects the conformational ensemble, the conformational selection model was developed. The conformational selection model postulates that in the absence of ligand, the ligand bound conformation exists in an equilibrium with the unbound conformation, though the ligand bound conformation may be at a higher energy and therefore less

populated *(10)*. As ligand binds, the conformational landscape is remodeled such that the equilibrium favors the ligand bound conformation. This model has experimental evidence whether the ligand is a small molecule, another protein, or DNA *(10)*.

As the ligand binds, the conformational shift leads to a change in protein dynamics. This change in dynamics upon ligand binding can be present regardless of the class of ligand. By monitoring the change/s in protein dynamics upon ligand binding, insights about the ligand bound conformation can be gained. If a ligand stabilizes a secondary structural element or if a rearrangement of the tertiary structure is made, this can impact protein function. The impact of changes in protein dynamics on protein function is explored in chapters 2-5.

Allostery

Ligand binding, or another energetic perturbation such as a mutation, can have effects on the protein structure and dynamics at areas distal from the interaction/perturbation site. This is called allostery *(11,12)*. In the original model of allostery (the MWC model or the symmetry model), an allosteric protein complex exists in one of two states: a high affinity state or a low affinity state (**Figure 1.3**). These two states exist in equilibrium with one another even when ligand is not present *(12)*. All subunits within the complex adopt the same conformation, thus leading to the alternative name for this model: the symmetry model. In the WMC model, the binding of the allosteric effector shifts the equilibrium such that the high affinity conformation is favored *(12,13)*.

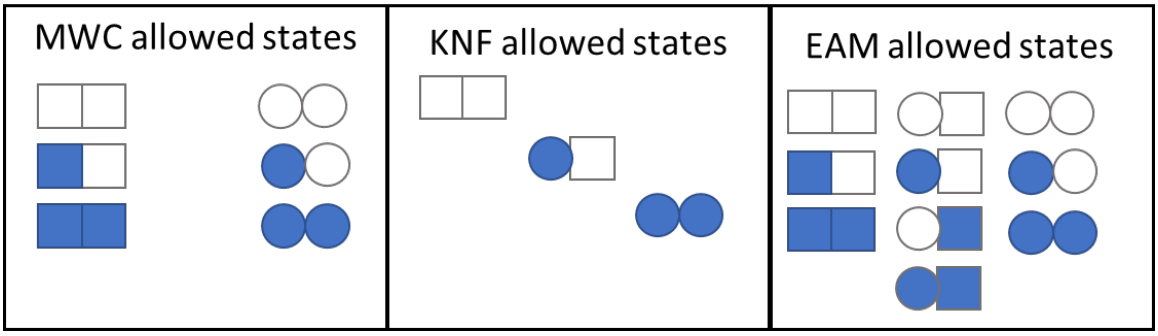

Figure 1.3 Allowed states for each of the proposed allosteric models. A homodimeric protein is used as an example. In each model squares represent the low affinity states and circles represent the high affinity states. Filled shapes represent ligand bound and open shapes represent protein not bound to ligand. The ensemble allosteric model (EAM) allows for more conformational heterogeneity than the other two proposed models. Figure adapted from Hilser *et al.* (2012).

However, the MWC model fails to explain negative cooperativity (upon binding of an allosteric effector or substrate, substrate binding is less favorable) *(13)*. This was addressed by a follow up model called the KNF model or the sequential model (**Figure 1.3**). In the KNF model there are still two states, the high affinity state and the low affinity state, but, unlike the MWC model, conformational change is a direct effect of ligand binding. This conformational change is communicated through the protein:protein interface such that bound subunits also change conformation. This model is not an ensemble model, instead it postulates that the bound conformation is only adopted in the presence of ligand (induced fit).

In order to reconcile the ensemble nature of proteins with the presence of negative cooperativity, the ensemble allosteric model (EAM) was developed *(12)*. In this model, proteins can be simplified into independent domains where each domain traverses the energy landscape sampling different conformations. These domains can exist in two states: the high affinity state and the low affinity state. As one domain adopts a high

affinity state, this information is communicated to the other domain/s causing a rearrangement of their energy landscape and the distribution of the conformation/s. Because each of these domains exist in an equilibrium of conformations, each possible combination of high and low affinity states can be accounted for by this model (**Figure 1.3**). This model is also not limited to just ligand binding, any energetic perturbation that can rearrange the energy landscape can be viewed as an allosteric effector. One of these potential perturbations is a point mutation. Investigation of a point mutation altering the energetic landscape is further explored in chapter 3. This model also allows for the investigation of allosteric networks that are not rooted in structural change.

Allostery can be present in a protein even in the absence of observed structural change *(14)*. In this case, it is a change in conformational entropy that is detected when a ligand binds. The mean conformation remains the same, but fluctuations about this mean are altered in amplitude or frequency. This is, in essence, the dynamics of the protein. Because most experimental methods track the average conformation of a protein, there may appear to be no change in conformation but the interconversion between the conformational ensemble is altered *(15)*. By tracking changes in dynamics upon mutation or ligand binding, the allosteric network within a protein can be probed. Chapter 3 of this thesis provides an example of a protein whose dynamics are altered upon a point mutation, but the structure of the protein remains the same. This chapter focuses on the capsid protein (Cp) of Hepatitis B virus (HBV). Upon mutating this one residue, the function of the Cp is abrogated (it is no longer capable of assembling into capsids) showing that this mutation has a functional effect on the protein, but the structure is

unaltered. An investigation into the changes in dynamics of the Cp upon mutation showed that regions of the protein distal from the mutation site were affected by the point mutation. This led to the conclusion that altering the dynamics in regions distal to the mutation site had a functional effect on the Cp without altering the mean conformation of the protein. Therefore, an allosteric network exists within the Cp.

This discussion of allostery leads to the question of whether or not every protein has the potential to possess an allosteric network. If these communication networks are present even in the absence of structural change, can they be observed in every protein? It would make sense that in the presence of an energetic perturbation (for example, ligand binding or mutation), the energetic landscape of the protein will be remodeled, and the conformational ensemble will be affected. Input of energy or alteration of the energy will change the strength and, therefore, the longevity of the noncovalent interactions that make up protein structure. Protein function is determined by the structure of the protein. If an allosteric network exists, the information in this network is transmitted through the formation, strength, and longevity of the noncovalent interactions that make up protein secondary, tertiary, and quaternary structure. Can a deeper understanding of these networks and their role in protein function give insight into how proteins work, be it a structural protein, and enzyme, a molecular motor, etc.? By understanding the effects of ligand binding or mutation on a protein, even at regions distal from the energy perturbation, the function of the protein may be better understood. Does this protein bind other proteins once a ligand is bound? Investigation into changes in protein dynamics may give insight into what this interaction signal is (See Chapter 4 for an example of

this). Does the binding of another protein alter the function of the studied protein? By studying this type of interaction, the mechanism behind the function of the protein can begin to be determined (see Chapter 5 for an example of this).

In this thesis, four model systems are tested for the presence of an allosteric network. To determine if there are changes to the protein upon either ligand binding or mutation, protein dynamics were probed through the use of hydrogen deuterium exchange mass spectrometry (HDX-MS). In each chapter, the changes in dynamics upon either ligand binding or mutation occur at the site of the energetic perturbation as well as in regions distal from the perturbation. The systems studied are a diverse representation of different protein complexes, all of which show allosteric effects.

Introduction to HDX-MS

The changes in protein dynamics upon an energetic perturbation can be probed through the use of hydrogen deuterium exchange mass spectrometry (HDX-MS). HDX was developed in 1954 by Linderstrøm-Lang to probe the hydrogen bonding network within proteins *(16)*. This was before the first structure of proteins was solved, so the hydrogen bonding within secondary structure had yet to be proven. It was discovered that short peptides show complete deuterium exchange rapidly while structured proteins can take days to show complete exchange *(17)*.

HDX relies on the fact that -NH, -OH, or -SH hydrogens will exchange with hydrogen in the solvent *(18)*. By replacing the bulk solvent with a deuterated buffer, an exchange of a hydrogen to a deuterium can occur. Sidechain -NH, -OH, and -SH groups will exchange rapidly while -NH groups in the protein backbone will exchange on the

milliseconds to days timescale *(17–19)*. The exchange of a hydrogen for a deuterium can be tracked through the use of mass spectrometry which is a sensitive enough analytical method to measure a net increase of one amu per exchange event.

The wide range of timescales observed for HDX of natively folded proteins is caused by the structured nature of the proteins. Amide hydrogens that are solvent accessible and not involved in secondary structural elements will exchange rapidly *(20)*. However, amide hydrogens that are involved in hydrogen bonding with the protein backbone (secondary structural interactions) or with other amino acid residue side chains (tertiary structural interactions) will exchange much more slowly. The ability of these protected amide hydrogens to exchange is reliant on local unfolding events *(21)*.

$$-NH\ (closed) \overset{k(open)}{\longleftrightarrow} -NH(open) \xrightarrow{k(exchange)} -ND(open) \overset{k(open)}{\longleftrightarrow} -ND(closed)$$

The exchange of these amide hydrogens (-NH) for a deuterium (-ND) is dependent on the rate of the unfolding event (k(open)). If the actual exchange of a hydrogen for a deuterium is rapid in comparison to the k(open) rate for a folded protein this is termed EX1 kinetics, if multiple unfolding events need to occur in order for an exchange event to occur then the kinetics of the exchange are termed EX2 *(18)*. EX1 can be used to view global unfolding/folding events while EX2 is commonly used to probe a structured protein's dynamics.

By tracking the exchange of hydrogen for deuterium, the longevity of the hydrogen bonding networks that make up protein structure can be probed. The more often a local unfolding event occurs, the faster the hydrogens in this region will be exchanged for a deuterium *(21)*. The local unfolding which allows for this exchange event plays a

role in protein dynamics. The more dynamic a region, the more deuterium will be taken up in this region at earlier time points *(19)*. If the hydrogen bonding network is stable, fewer local unfolding events will occur during the time course, and fewer deuterium will be taken up by the region *(20)*. Thus, in cases when less deuterium is taken up, the protein structure is more stable and less dynamic. A typical HDX-MS experiment is outlined in Figure 1.4.

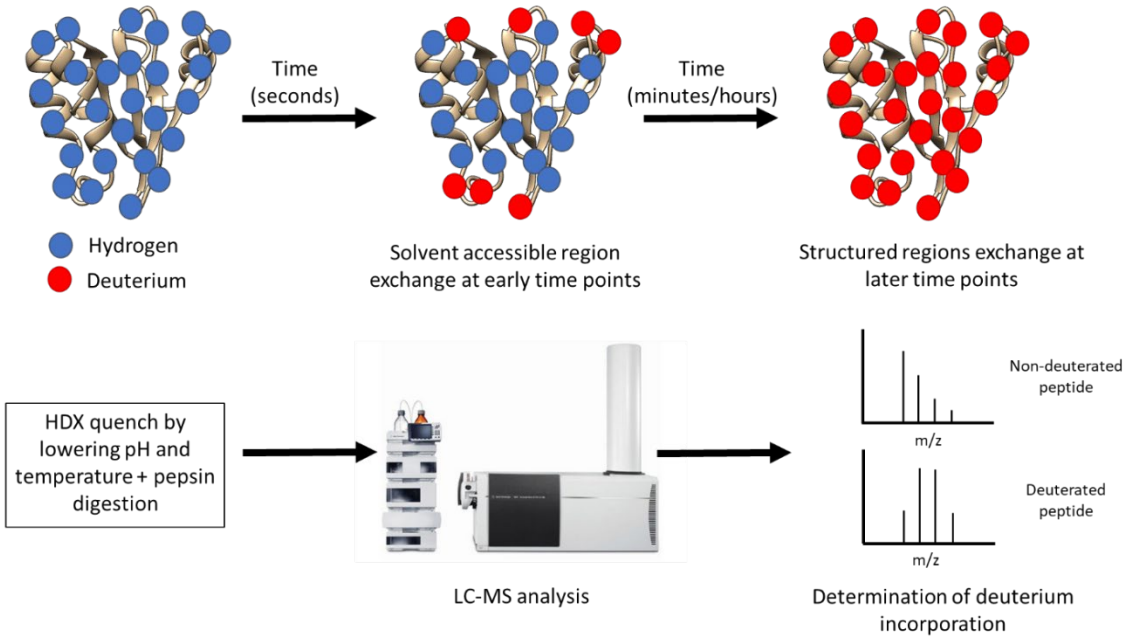

Figure 1.4 A typical HDX-MS workflow. The HDX reaction is initiated by diluting the protein 1:10 (v/v) into deuterated buffer. Solvent accessible regions exchange deuterium at early time points while more structured regions exchange deuterium at later time points. The reaction is quenched at various time points by lowering the pH and the temperature. The protein is then digested by a non-specific protease that has activity at low pHs (pepsin). The peptide mixture is then stored in liquid nitrogen until LC-MS analysis. LC-MS allows for the determination of how many deuterium are taken up per peptide through the analysis of the isotopic distributions for each peptide.

By comparing the amount of deuterium taken up by a complex either unbound or bound to a ligand, the changes in protein dynamics and solvent accessibility upon ligand

binding can be determined *(22)*. At early time points, protection from HDX in the ligand bound protein/complex allows for the determination of the binding site of the ligand, if there are changes in solvent accessibility in this region *(20)*. At later time points, changes in deuterium exchange allow for the determination of changes in protein dynamics upon ligand binding. These changes in dynamics may occur even if the structure of the protein/protein complex remains similar. By comparing the two conditions (bound and unbound) determinations about changes in the conformational ensemble can be made *(22)*.

Applying HDX-MS to Track Allosteric Networks

HDX-MS was used to investigate changes in protein dynamics upon ligand binding in four systems. These experiments involved tracking the changes in deuterium exchange upon DNA binding or small protein binding as well as changes in dynamics upon introduction of a point mutation. In each of the systems studied, there were observed changes in deuterium uptake upon energetic perturbation in regions distal from the perturbation site. This implies that allostery exists in each of the studied systems. The role of allostery in the function of each of the protein complexes studied was investigated to develop a deeper understanding of how these protein complexes function. In the replication protein A (RPA) system (Chapter 2), HDX-MS was used to determine the effects on the conformation of RPA upon ssDNA binding. This provided evidence to support an existing model that RPA has a more dynamic half and a less dynamic half. Within the heterotrimeric protein complex, domains that are distal from the ssDNA interaction sites showed changes in deuterium exchange. These domains are involved in

protein:protein interactions, showing that upon ssDNA binding, the allosteric network communicates this binding event to functional regions of the complex which are spatially distant from the binding site. In Chapter 3, the capsid protein homodimer (Cp) of hepatitis B virus (HBV) was used as a model system to study the changes in protein dynamics upon mutation of a single residue. This mutation prevents the viral capsid from assembling and allowed for the determination of key protein regions that must remain stable for capsid assembly to occur productively. The changes in protein dynamics observed within the Cp occur throughout the protein structure. There was a significant difference in deuterium exchange located at the intra-dimer interface which is spatially distal from the dimer:dimer interface where the mutation occurs. This data showed that there is an allosteric network within the Cp that connects the two interaction interfaces and alteration of the dynamics within this network has a direct effect on protein function. In Chapter 4, the difference in deuterium exchange when dsDNA is bound to the Type IE CRISPR interference complex Cascade allowed for the functional prediction of the binding site of the non-target strand of DNA as well as revealing a key Cas3 recruitment site. This Cas3 recruitment site was located in a region that is not directly involved with dsDNA recognition or binding. This region is on the surface of the Cse1 protein within the tail of the Cascade complex, and changes in protein dynamics in this region serve as a recruitment signal for the helicase/endonuclease Cas3. This illustrates the functional use of an allosteric network to transmit signals between two distal regions of a protein or protein complex. Finally, in Chapter 5, the changes in the conformational landscape of the Type IF CRISPR interference complex, Csy, upon binding of two different anti-

CRISPRs were determined using HDX-MS as a guide. The changes in protein dynamics spanned throughout the Csy complex upon binding of either anti-CRISPR, and upon binding the function of the Csy complex is negated. These experiments show that the function of a whole protein complex can be obstructed by the binding of a small protein because this binding event has long reaching effects on protein structure and/or dynamics. These long-range effects are a perfect example of allostery and show that these networks can be used by invaders or inhibitors to completely negate a protein's function. In this thesis, the following chapters illustrate the presence of allosteric networks across a wide range of protein complexes. These complexes have various functions and, in each case, the allosteric network which is present provided important insight into that function.

References

(1) Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys.* **1968**, *65* (1), 44–45. https://doi.org/10.1051/jcp/1968650044.

(2) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (45), 15873–15880. https://doi.org/10.1073/pnas.1411798111.

(3) Ben-Nissan, G.; Sharon, M. Capturing Protein Structural Kinetics by Mass Spectrometry. *Chem. Soc. Rev.* **2011**, *40* (7), 3627–3637. https://doi.org/10.1039/c1cs15052a.

(4) Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8* (6), 1181–1190. https://doi.org/10.1110/ps.8.6.1181.

(5) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science (80-.).* **1995**, *267* (5204), 1619–1620. https://doi.org/10.1126/science.7886447.

(6) Oliveberg, M.; Wolynes, P. G. The Experimental Survey of Protein-Folding Energy Landscapes. *Q. Rev. Biophys.* **2005**, *38* (3), 245–288.

https://doi.org/10.1017/S0033583506004185.

(7) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339. https://doi.org/10.1038/nature13001.

(8) Wiederschain, G. Y. *Membrane Protein Structure Determination. Methods and Protocols*; 2011; Vol. 76. https://doi.org/10.1134/s0006297911110113.

(9) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **1958**, *44* (2), 98–104. https://doi.org/10.1073/pnas.44.2.98.

(10) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796. https://doi.org/10.1038/nchembio.232.

(11) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(12) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41* (1), 585–609. https://doi.org/10.1146/annurev-biophys-050511-102319.

(13) Koshland, D. E.; Hamadani, K. Proteomics and Models for Enzyme Cooperativity. *J. Biol. Chem.* **2002**, *277* (49), 46841–46844. https://doi.org/10.1074/jbc.R200014200.

(14) Cooper, A.; Dryden, D. T. F. Allostery without Conformational Change - A Plausible Model. *Eur. Biophys. J.* **1984**, *11* (2), 103–109. https://doi.org/10.1007/BF00276625.

(15) Nussinov, R.; Tsai, C. J. Allostery without a Conformational Change? Revisiting the Paradigm. *Curr. Opin. Struct. Biol.* **2015**, *30*, 17–24. https://doi.org/10.1016/j.sbi.2014.11.005.

(16) Baldwin, R. L. Early Days of Protein Hydrogen Exchange: 1954-1972. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (7), 2021–2026. https://doi.org/10.1002/prot.23039.

(17) Hvidt, A.; Linderstrøm-Lang, K. Exchange of Hydrogen Atoms in Insulin with Deuterium Atoms in Aqueous Solutions. *BBA - Biochim. Biophys. Acta* **1954**, *14* (C), 574–575. https://doi.org/10.1016/0006-3002(54)90241-3.

(18) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for

Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/C0CS00113A.

(19) Al-Naqshabandi, M. A.; Weis, D. D. Quantifying Protection in Disordered Proteins Using Millisecond Hydrogen Exchange-Mass Spectrometry and Peptic Reference Peptides. *Biochemistry* **2017**, *56* (31), 4064–4072. https://doi.org/10.1021/acs.biochem.6b01312.

(20) McAllister, R. G.; Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **2015**, *54* (16), 2683–2692. https://doi.org/10.1021/acs.biochem.5b00215.

(21) Jaswal, S. S. Biological Insights from Hydrogen Exchange Mass Spectrometry. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834* (6), 1188–1201. https://doi.org/10.1016/j.bbapap.2012.10.011.

(22) Sowole, M. A.; Konermann, L. Effects of Protein-Ligand Interactions on Hydrogen/Deuterium Exchange Kinetics: Canonical and Noncanonical Scenarios. *Anal. Chem.* **2014**, *86* (13), 6715–6722. https://doi.org/10.1021/ac501849n.

CHAPTER TWO

HYDROGEN-DEUTERIUM EXCHANGE REVEALS A DYNAMIC DNA-BINDING MAP OF REPLICATION PROTEIN A

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Faiz Ahmad

Contributions: Contributed to the understanding and application of the HDX-MS data presented in the manuscript. Wrote and edited the manuscript.

Author: Angela Patterson

Contributions: Developed and performed HDX-MS experiments. Participated in writing and editing the manuscript

Author: Jaigeeth Deveryshetty

Contributions: Contributed to the understanding of the application of the HDX-MS data presented in the manuscript. Wrote and edited the manuscript.

Co-Author: Jenna Mattice

Contributions: Participated in performing the HDX-MS experiment and data analysis. Reviewed and edited the manuscript.

Co-Author: Nilisha Pokhrel

Contributions: Contributed to the understanding and application of the HDX-MS data presented in the manuscript. Reviewed and edited the manuscript.

Co-Author: Brian Bothner

Contributions: Supervised the HDX-MS experiment. Reviewed and edited the manuscript.

Co-Author: Edwin Antony

Contributions: Supervised the protein purification for the HDX-MS experiment. Supervised the application of the HDX-MS data to RPA system. Wrote and edited the manuscript.

Manuscript Information

Faiz Ahmad, Angela Patterson, Jaigeeth Deveryshetty, Jenna R. Mattice, Nilisha Pokhrel, Brian Bothner, Edwin Antony

Nucleic Acids Research

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__x_ Published in a peer-reviewed journal

Nucleic Acids Research
Volume 9, Issue 3
DOI: 10.1093/nar/gkaa1288

Abstract

Replication protein A (RPA) binds to single-stranded DNA (ssDNA) and interacts with over three dozen enzymes and serves as a recruitment hub to coordinate most DNA metabolic processes. RPA binds ssDNA utilizing multiple oligosaccharide/ oligonucleotide binding domains and based on their individual DNA binding affinities are classified as high versus low-affinity DNA-binding domains (DBDs). However, recent evidence suggests that the DNA-binding dynamics of DBDs better define their roles. Utilizing hydrogen–deuterium exchange mass spectrometry (HDX-MS), we assessed the ssDNA-driven dynamics of the individual domains of human RPA. As expected, ssDNA binding shows HDX changes in DBDs A, B, C, D and E. However, DBD-A and DBD-B are dynamic and do not show robust DNA-dependent protection. DBD-C displays the most extensive changes in HDX, suggesting a major role in stabilizing RPA on ssDNA. Slower allosteric changes transpire in the protein–protein interaction domains and linker regions, and thus do not directly interact with ssDNA. Within a dynamics-based model for RPA, we propose that DBD-A and -B act as the dynamic half and DBD-C, -D and -E function as the less-dynamic half. Thus, segments of ssDNA buried under the dynamic half are likely more readily accessible to RPA-interacting proteins.

Introduction

In eukaryotes, the single-strand DNA (ssDNA)-binding protein replication protein A (RPA) is essential for most DNA metabolic processes including DNA replication,

repair, and recombination. RPA binds to ssDNA with high affinity (KD $<10^{-10}$M) *(23)* and protects it from degradation by exo- and endonucleases. Formation of RPA-ssDNA complexes trigger the ATM/ATR cellular DNA damage checkpoint response *(24,25)*. RPA physically interacts with over three dozen DNA processing enzymes and recruits several of them onto DNA *(26)*. Finally, RPA hands-off the DNA to these enzymes and correctly positions them to facilitate their respective catalytic activity *(27)*. Plasticity for such multiplexed functional roles for RPA can be ascribed to its flexible structure and its corresponding context-dependent interactions with DNA *(28,29)*.

RPA is a heterotrimer composed of RPA70, RPA32, and RPA14 subunits. Structurally, it can be further partitioned into six oligonucleotide/oligosaccharide-binding folds (OB-folds; denoted A-F in **Figure 2.1a**). RPA70, the largest subunit, is composed of OB-folds F, A, B, and C. RPA32 consists of an N-terminal phosphorylation region, OB-fold D, and a C-terminal winged-helix domain (wh). RPA14 harbors OB-fold E (**Figure 2.1a**). The heterotrimer is constitutively held together by the trimerization core (Tri-C) through interactions between domains C, D and E (**Figure 2.1a**) *(30–32)*. RPA mediates protein-protein interactions with over three dozen proteins primarily through two dedicated protein-interaction domains (PIDs) *(33–35)*. For clarity, here we partition the OB-folds as DNA binding domains (DBDs) and protein interaction domains (PIDs) based on their established biochemical functions. Domains A, B, C, D and E are DBDs whereas F and wh are PIDs. The F-domain in RPA70 (PID^{70N}) and the 'wh' domain in RPA32 (PID^{32C}) mediate direct physical interactions to most RPA interacting proteins (RIPs). DBDs and PIDs are connected by flexible linkers of varying lengths and thus

allow RPA to adopt a multitude of configurations on ssDNA *(28,36,37)*. The DNA-bound crystal structure of Ustilago maydis RPA shows how the DBDs of RPA are arranged on ssDNA (16). DBDs-A, B, C and D engage the DNA and parlay 5′-3′ directionality with DBD-A residing towards the 5′ end *(32,38–42)*. PID^{70N} and PID^{32C} are not present in this structure, and DBD-E does not contact the DNA. Most of the intervening linker regions are also disordered (i.e. flexible; electron density not visualized) in the structure.

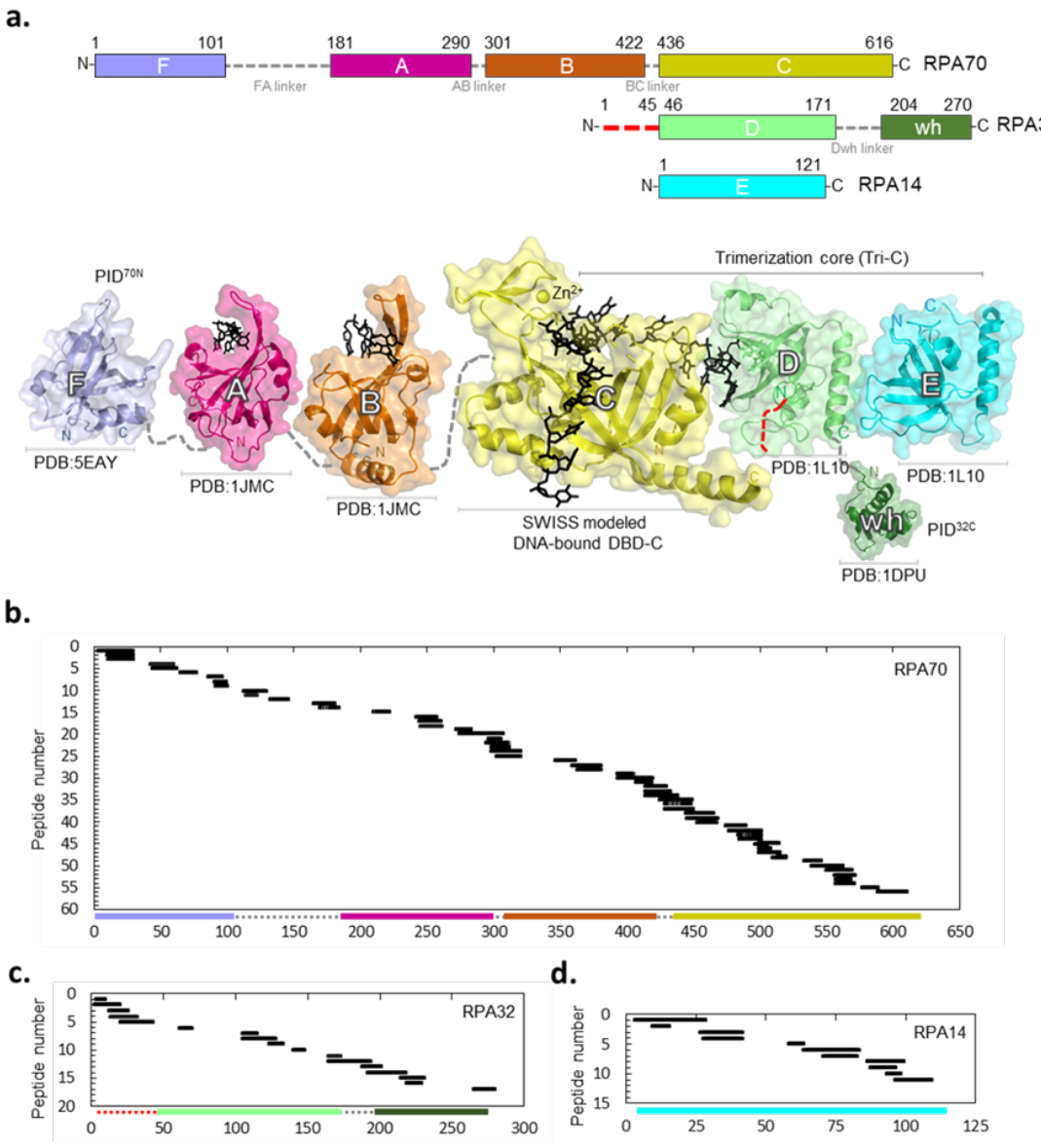

Figure 2.1. Sequence coverage of the DNA binding and protein interaction domains of RPA. a. Top - Domain composition of human RPA70, RPA32, and RPA14 subunits and their respective linkers. The red dotted line in the N-terminus of RPA32 denotes the region of phosphorylation. Bottom – The crystal structures of the individual domains of RPA are depicted and color coded to match the domains on the top panel. For domain C, DNA is modelled onto DBD-C using SWISS-MODEL with the *Ustilago maydis* structure as reference. Peptides identified in the MS analysis corresponding to b. RPA70, c. RPA32, and d. RPA14 are shown. 56 peptides from RPA70 (59% sequence coverage), 17 peptides from RPA32 (50% sequence coverage), and 11 peptides from RPA14 (49% sequence coverage) were identified. The y-axis indicates the peptide fragment number, and the x-axis denotes peptides fragments arranged in order of increasing residue numbers from the N-terminus of each subunit as denoted in Supplemental Table 1

Although RPA binds to ssDNA with high affinity, it must be removed by RIPs that bind ssDNA with much lower affinity. This is achieved by the differential DNA binding properties of individual DBDs *(43)*. In solution, human RPA displays multiple NaCl concentration-dependent DNA binding modes denoted by the occluded site-size *(44)*. This varies from 22 nucleotides (nts) in conditions below 50 mM NaCl to 28-30 nts at higher salt concentrations of up to 1.5 M NaCl. These binding modes are propagated by selective binding of one or more DBDs. Furthermore, transitions between the DNA binding modes likely drive the diffusion and remodeling of RPA on ssDNA *(45)*. Thus, multiple configurations of RPA on ssDNA can be formed by using one or a subset of the DBDs. Two models have been proposed for the utilization of DBDs: A sequential DBD binding model posits that smaller occluded site-sizes are likely due to the higher affinity ssDNA binding properties of DBDs-A and B which bind first onto DNA *(46)*. Subsequent binding of the other DBDs in the trimerization core drive the 30 nt mode. An alternate dynamic DBD engagement model is built on the intrinsic on-off properties of each DBD with no particular emphasis on the order of DBD binding *(43)*. Both models are built on decades of DNA binding studies of isolated DBDs, where DBDs-A and B are

classically defined as the high affinity binders with DBDs-C and D as low affinity ssDNA binding domains *(47–49)*. A few studies have also reported DBD-E and PID^{70N} to possess very weak DNA binding properties *(50,51)*. In both scenarios, a canonical bias exists that DBDs-A and -B are drivers of stable or robust complex formation between RPA and ssDNA. Therefore, many models also invoke that RIPs will have to remove the interactions between DBDs-A & -B and DNA to displace or remodel RPA.

Recent cryoEM and single molecule studies present a more nuanced view of the DBDs of RPA *(43,52,53)*. CryoEM structures of Saccharomyces cerevisiae RPA on longer ssDNA (100 nts) reveals an interaction between the DBDs when multiple RPA molecules are bound *(53)*. In these structures, the resolution of DBDs-A and -B are poor compared to Tri-C. Furthermore, DBD-A from one RPA molecule physically interacts with DBD-E of the neighboring RPA molecule. Such an arrangement opens up segments of ssDNA bound between two RPA molecules, likely acting as DNA binding regions for RIPs (32). In the single-molecule experiments, DBD-A and DBD-D have four binding states on DNA with the lifetime of each state differing between the two DBDs *(43,52)*. Associated bulk fluorescence kinetic experiments have shown conditional displacement of DBD-A by Tri-C *(52)*. These findings contradict the high-affinity binding label ascribed to DBDs-A and B as these domains are either not bound to DNA, or readily outcompeted by Tri-C. The dynamics and arrangement of individual DBDs can also be altered by post-translational modifications and interactions with RIPs. Phosphorylation at S178 in S. cerevisiae RPA70, a position on the FA-linker adjacent to DBD-A, alters its positioning with respect to the neighboring DBDs *(53)*. Similarly, the presence of Rad52,

a mediator protein that functions in homologous recombination, selectively alters the number of states formed by DBD-D on ssDNA *(52)*. This alteration enables Rad52 to load onto the 3′ end of the RPA-coated DNA by remodeling DBD-D. In contrast to classical DNA-binding studies of RPA, these new tools provide information on individual DBDs whilst investigating full-length RPA. Nevertheless, information on the DNA-binding dynamics of DBD-B, C, E and the PIDs are lacking due to technical limitations.

Here, we use time-resolved hydrogen-deuterium exchange mass spectrometry (HDX-MS) and full-length human RPA to investigate the protein solvent accessibility and kinetic stability of each DBD on ssDNA. HDX-MS is ideal for studying protein folding and its conformational dynamics *(54–57)*. HDX-MS measures the rate of amide-H exchange to report on changes in the local environment. D_2O exchange rates are shaped by dynamics, H-bonding, secondary structure, and solvent exposure. Increased rates of D_2O exchange result from greater solvent accessibility, disruption in backbone H-bonding, or increased dynamics. Decreased D_2O exchange rates suggest less solvent accessibility or local stabilization (i.e. H-bonding and/or formation of secondary structure). As such, exchange rate disruptions are valuable indicators of protein allostery, dynamics and interfaces *(58)*. In recent years, HDX-MS has gained attention as a powerful tool to study protein-DNA binding interactions *(54,57,59,60)*. Here, using HDX-MS we identify the effects on solvent accessibility, which are interpreted in terms of protein dynamics caused by changes in the contacts between the various DBDs, linkers, and PIDs of human RPA upon complex formation with ssDNA. Based on the data, in the context of the dynamics-based model for RPA, we propose that RPA can be

envisioned to function as two halves: DBD-A, DBD-B, and PID^{70N} (all in RPA70) serve as the dynamic half of RPA. Tri-C, consisting of DBD-C (RPA70), DBD-D (RPA32) and DBD-E/RPA14, functions as the less dynamic half with DBD-C serving as the anchor that likely controls the residence time of RPA on ssDNA.

Material and Methods

HDX-MS

Human RPA was purified as described *(44)*. RPA and DNA (dT)30 were mixed as 1:1 molar equivalents (3.9 mg/mL RPA) and diluted 1:10 into deuterated reaction buffer (30 mM HEPES, 200 mM KCl pD 7.8 in D_2O). Undeuterated controls (zero time points) were diluted 1:10 into non-deuterated buffer. 10 µL of the reaction was removed at each time point (0.008, 0.05, 0.5, 3, and 24h) and quenched on ice by adding 60 µL of 0.75% formic acid (FA, Sigma) pH 2.5 containing 0.25 mg/mL porcine pepsin (Sigma). Each sample was digested for two minutes with vortexing every 30 seconds and then flash frozen in liquid nitrogen. Samples were stored in liquid nitrogen until liquid chromatography mass spectrometry (LC-MS) analysis. All reactions were performed in triplicate.

LC-MS analysis of RPA was performed on a 1290 UPLC series chromatography stack (Agilent Technologies) coupled directly to a 6538 UHD Accurate-Mass QTOF LC/MS mass spectrometer (Agilent Technologies). Before electrospray−time-of-flight (ESI-TOF) analysis, peptides were separated on a reverse phase (RP) column (Phenomenex Onyx Monolithic C18 column, 100 × 2 mm) at 1 °C using a flow rate of 500 µL/min under the following conditions: 1.0 min, 5% B; 1.0−9.0 min, 5−45% B;

9.0−11.8 min, 45-95% B; 11.80−12.0 min, 5% B; solvent A = 0.1% FA in water (Thermo-Fisher) and solvent B = 0.1% FA in acetonitrile (ACN, Thermo-Fisher). Data were acquired at 2 Hz over the scan range 50−1700 m/z in positive mode. Electrospray settings were as follows: nebulizer set to 3.7 bar, drying gas at 8.0 L/min, drying temperature at 350 °C, and capillary voltage at 3.5 kV.

Peptides were identified as outlined in Berry et al. *(55)* using MassHunter Qualitative Analysis version 6.0 (Agilent Technologies), Peptide Analysis Worksheet (PAWs, ProteoMetrics LLC), and Peptide Shaker version 1.16.42 paired with Search GUI version 3.3.16 (Compomics). Deuterium uptake was determined using HDExaminer version 2.5.1 (Sierra Analytics).

Data Analysis

Average relative deuterium uptake difference (ΔHDX) averaged over all HDX incubation periods was calculated using Equation 1 *(60)*.

$$\Delta\text{HDX} = \sum_i \frac{A(t_i) - B(t_i)}{A(t_i)}$$

The ΔHDX between bound and free RPA was calculated with the above equation, where A is the deuterium uptake for sample A (RPA+DNA) at one time point (ti), and B is the deuterium uptake for sample B (free RPA) at the same time point (ti). Summation of ΔHDX at all five time points together gives ΔHDX for a peptide (Figure 2). Mean of the deuterium uptake value of a peptide was used to generate a heatmap and plotted using GraphPad Prism 8.4.1. Crystal structures were rendered using PyMOL (Schrödinger, LLC). The uptake curve for each peptide was plotted using a non-liner/non-logarithmic scale using MS-Excel. Each data point represents the mean value, and error bars

represent the ±SD of mean for three biological replicates. Peptides with a SD> 0.5 were excluded from this study. To compare the difference at time points with ±DNA statistical significance were determined. *, the first incubation time point that shows statistical difference is indicated in the linear plot. No significance is indicated as 'ns' and $p<0.05$ was considered statistically significant (* p-value < 0.05, ** $p<0.01$, ***$p<0.001$, two-tailed, unpaired t test). In the full-length RPA plasmid, RPA32 is his-tagged, thus HDX peptide 17 has an extra 6 His residues at its C-terminus i.e., TDAEHHHHHH.

Results

HDX-MS Generates a Conformational Landscape for DNA-Induced Changes in Human RPA

To determine the ssDNA binding associated conformational changes in RPA, we performed HDX-MS experiments with human RPA in the absence and presence of ssDNA. A (dT)30 oligonucleotide was used to provide adequate binding sites for all the domains of RPA as the binding site-size for RPA is ~22-28 nt *(44)*. Since RPA binds to ssDNA stoichiometrically, we used an equimolar concentration of RPA and (dT)30. Although RPA is a large (1007 amino acids), we observe excellent sequence coverage (**Figure 2.1b-1d**). We identified 56, 17, and 11 different peptides for the three subunits, respectively. This provided an overall 51% sequence coverage (RPA70 – 363 of 616 aa ~59%, RPA32 – 139 of 276 aa ~50%, RPA14 – 59 of 121 aa ~49%) with the majority of the unreported sequence in the FA-linker (**Figure 2.1b**). For much of the complex, multiple overlapping peptides were identified. These were grouped and assigned Roman

numerals for clarity (**Figure 2.2**). There are 14, 8, and 5 groups for three RPA subunits, respectively (**Figures 2.2a, 2.2b and 2.2c**).

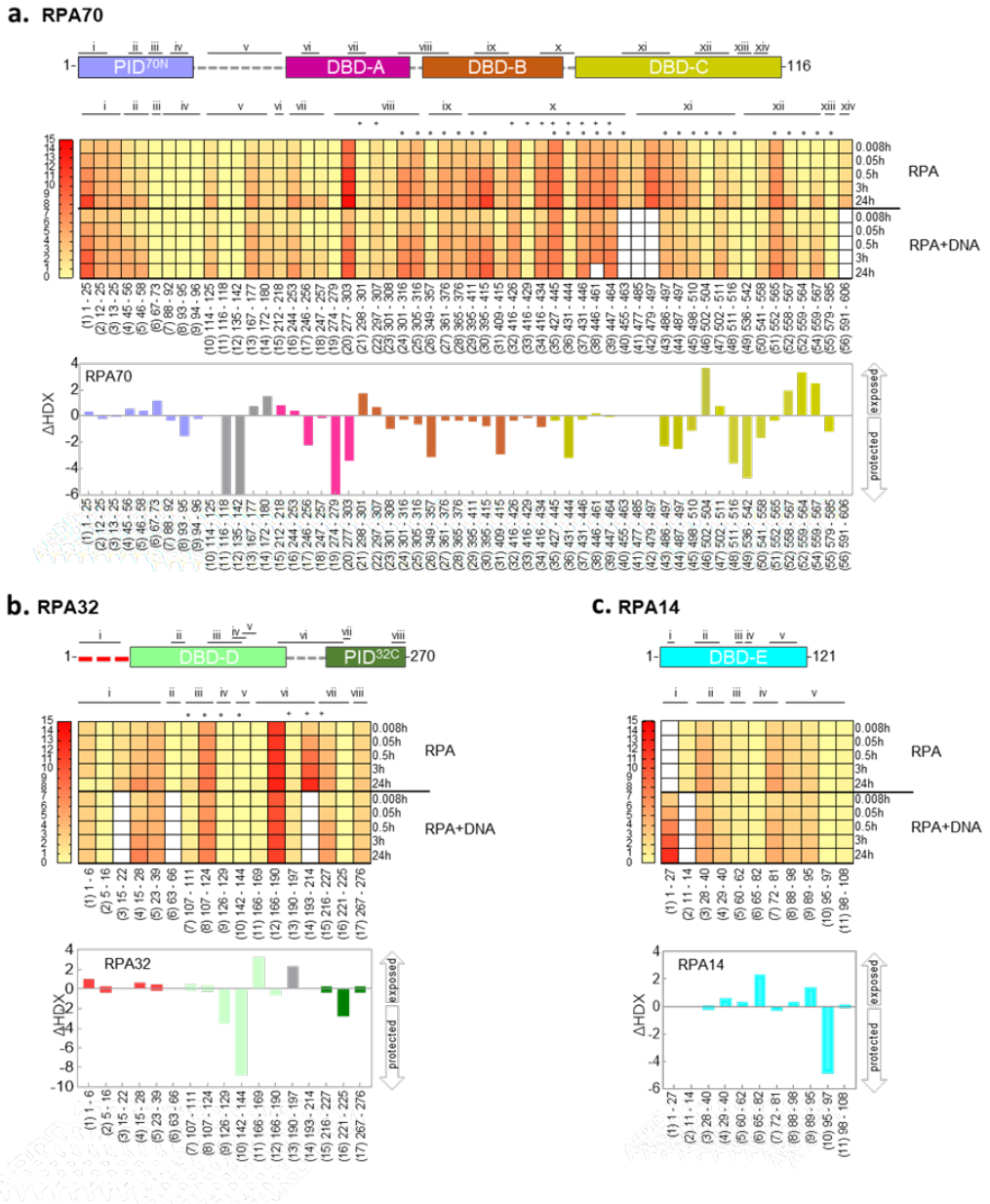

Figure 2.2. Global deuterium uptake profiles and DNA induced changes in RPA. Heat map representing the deuterium content of a. RPA70, b. RPA32, and c. RPA14 peptides in unbound and DNA bound states as a function of time. Each row represents a timepoint and each column represents a peptide. The heat map gradients are color coded yellow (low deuterium uptake) to red (high deuterium uptake). The white boxes depict absence

of a corresponding dataset for specific peptides under one condition (presence or absence of DNA). • Denotes peptides in DNA binding sites, and + marks peptides from the linker regions. To map the peptide location in RPA70, RPA32 and RPA14, peptides in same region were grouped and assigned roman numerals. Binding to ssDNA leads to varying patterns of protection by HDX-MS and the differences are denoted as bar graphs below the respective heat maps. The arrows denote whether the peptides were protected or exposed to solvent upon DNA binding. The data show the cumulative difference in deuterium uptake (ΔHDX) in RPA over all five HDX timepoints

RPA binds rapidly to ssDNA (k_{on} =1.1±0.6x108 $M^{-1}s^{-1}$) *(52)*. Therefore, we reasoned direct DNA-binding induced changes would occur rapidly, while diffusion and DBD rearrangement associated deuterium exchange would occur on a longer timescale. To get a direct readout of the kinetics of deuterium exchange in the absence and presence of ssDNA we performed the experiments as a function of time. The data are described in terms of the kinetics and patterns of HDX. In terms of the HDX kinetics, two prominent deuterium exchange patterns are observed which we classify as fast and slow exchange. Fast exchanges are deuteration that has occurred before the first time point captured in our experiments. RPA binding to ssDNA is diffusion limited; thus, the initial DNA-induced rapid conformational changes occur faster than our ability to capture them. Differences in exchange kinetics that occur only on a fast timescale, suggest both rapid and stable changes on binding to ssDNA. If exchange is only observed at early time points, and the amount of deuterium taken up by the peptide does not increase over time, the exchange is consistent with lasting H-bond networks and a stable structure. Slow exchange is used here to describe deuterium exchange occurring at later time points. Slow exchange is a result of dynamic H-bond networks leading to solvent exposure. This infers slower on/off rates due to conformational changes and dynamics. Difference in HDX rates that occur over a 3 to 30 min interval are likely due to rearrangements of the

domains' organization and other allosteric effects, which we have classified as slow exchange.

With respect to HDX patterns, we either observe a protection or exposure of certain regions of RPA upon binding to ssDNA. In our experiments, data is collected after mixing RPA with ssDNA. Since the binding is diffusion limited, a change in uptake at the first time point (30 sec) likely reflects a direct DNA-induced change. If a domain forms a long-lived stable complex on DNA, a corresponding long-lived difference in deuterium exchange for a particular peptide will be observed compared with RPA alone. We classify such a signal as protection. This can be thought of as protein surfaces that become buried upon DNA binding and do not come apart. If exchange continues at the minutes to hours timepoints, then those peptide regions are transiently exposed to solvent and are therefore 'less' protected, hence more exposed or dynamic.

Data for all the peptides are summarized as a heat map where red and yellow indicate high and low deuterium uptake, respectively (**Figures 2.2a-c**). White boxes denote absence of data as one of the corresponding peptides were not identified in the MS analysis. Peptides 1 to 5, and 10 to 56 show the highest degree of deuteration and correspond to PID^{70N}, DBDs -A, -B, and -C in RPA70, respectively (**Figure 2.2a**). For RPA32, peptides 7 to 8 and 11 to 14 in DBD-D show larger changes in deuteration when bound to DNA. Peptides 15 and 17, which are in the winged-helix domain (PID^{32C}) show less exchange (**Figure 2.2b**). Thus, both protein interaction domains in RPA (PID^{70N} and PID^{32C}) show minimal changes that transpire on a slower timescale, likely reflecting conformational changes driven indirectly after DNA binding to the DBDs. Peptides 5, 6

and 8 to 11 in DBD-E show moderate deuterium exchange, but the fast exchange kinetics coincide with the rate of DNA binding (**Figure 2.2c**). This is consistent with interactions between DBD-E and DNA as suggested by crosslinking experiments *(50)*.

To construct a global landscape of the DNA induced deuterium exchange in RPA, we quantified the average relative deuterium uptake difference (ΔHDX) in the absence and presence of DNA (**Figure 2.2**). ΔHDX for each DNA binding experiment relative to apo RPA is averaged over time for every peptide. The scale of the bars denotes the net change in deuterium uptake. A change in the upward direction denotes increased deuterium uptake ('less' protected or exposed) in the presence of DNA. Conversely, change in the downward direction denotes less uptake (protected) upon DNA binding. ΔHDX shows robust deuterium exchange in DBD-C with additional changes in DBD-A and DBD-B as expected from known DBD-ssDNA interactions (**Figure 2.2a**). Similarly, ΔHDX are observed for DBD-D (RPA32; **Figure 2.2b**) and DBD-E (**Figure 2.2c**). DBD-E binding to DNA has been observed in a few biochemical studies, but not in the X-ray or cryoEM structures of RPA-DNA complexes. Our data show DNA-driven fast exchanges in DBD-E and supports interactions with DNA.

To best represent the ΔHDX data, we mapped the peptides identified in the MS analysis onto the available crystal structures of human RPA (**Figures 2.3, 2.4, 2.5**). Since a structure of full-length human RPA is not available, we constructed one for reference using the information available for the two halves of human RPA. We used the ssDNA-bound structure of DBD-A and -B (PDB ID:1JMC *(61)*) and RPA Tri-C (PDB ID: 1L10 *(31)*). However, the Tri-C structure is not in complex with DNA. Thus, we used the

DNA-bound structure of Ustilago maydis RPA as reference and modeled DNA onto human RPA Tri-C (1L10) using SWISS-MODEL *(62)*. Our data below are presented from the perspective of two dynamic halves in RPA: PID^{70N}, DBD A and DBD-B constituting the 'FAB half', and the 'Tri-C half' made of DBD-C, DBD-D, PID^{32C} and DBD-E.

The FAB Half of RPA Shows Highly Dynamic ss-DNA-Dependent Changes in HDX

PID^{70N}, DBD-A and DBD-B reside in RPA70 and are connected by flexible linkers. The 79 aa FA-linker connects DBD-A to the PID^{70N} and the shorter 10 aa long AB-linker links DBD-A and -B (**Figure 2.3a**). DBD-A and -B are well-established DNA binding domains (KD = 2 μM DBD-A; 20 μM DBD-B; 50 nM DBD-A+B) *(43,48,49)*. Canonically, these DBDs are considered as the high affinity DNA binding domains of RPA. For PID^{70N} we observe changes on longer time scales accompanied by a slow decrease in protection upon DNA binding (**Figures 2.3b-e**). Peptides 4 and 5 are part of the basic cleft of the proposed weak DNA binding site of PID^{70N}. Both peptides show similar total exchange with or without DNA, but the uptake rate is slower without DNA (**Figure 2.3c**). Peptide 6 is protected without DNA, becoming more dynamic (exchange over a long timescale), with slow conformational changes when DNA bound (**Figure 2.3d**). These observations suggest that the conformational changes in PID^{70N} in the presence of DNA occur on a timescale well past the initial formation of the RPA-DNA complex. Thus, PID^{70N} does not likely interact directly with DNA.

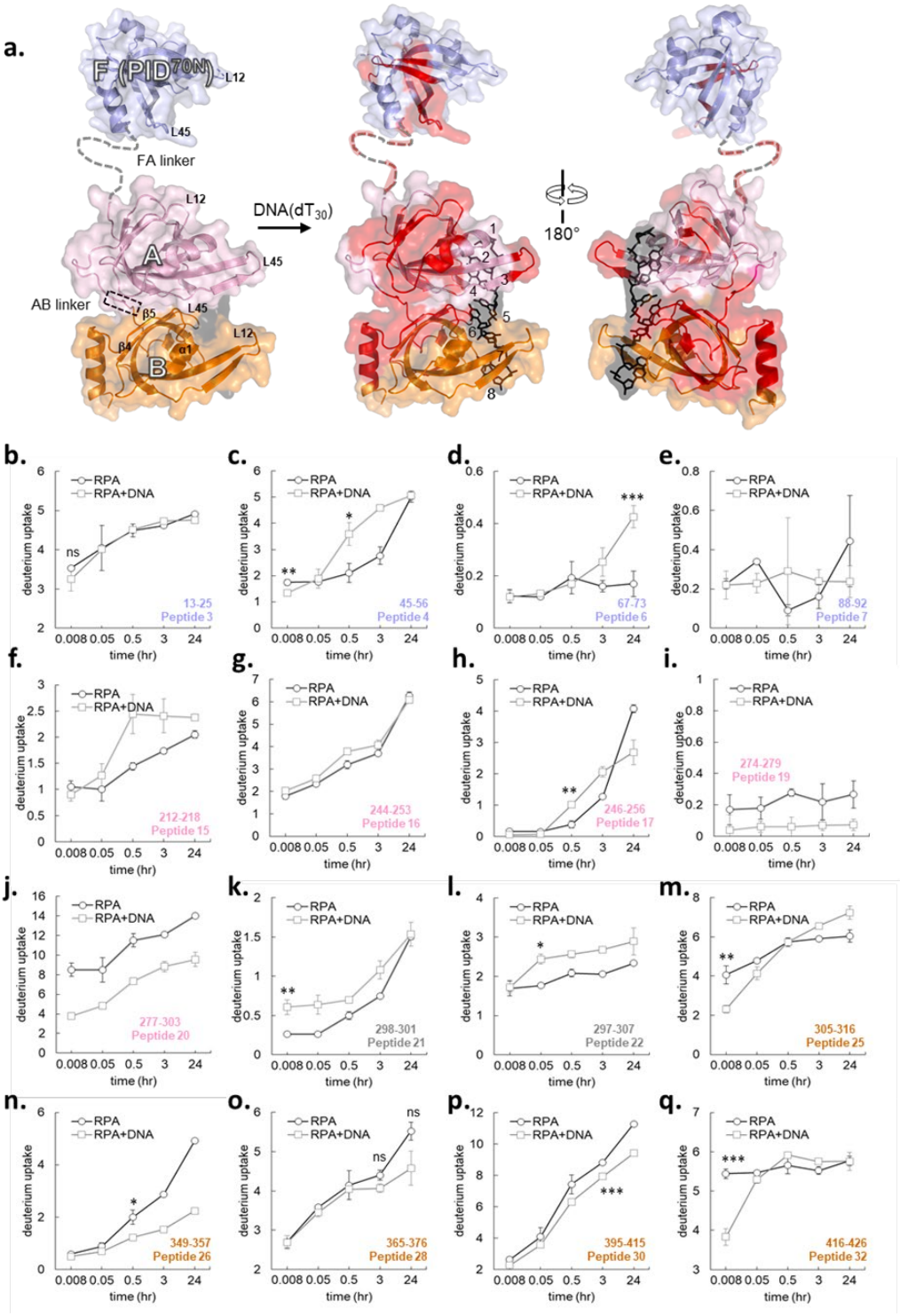
a.
F (PID 70N)
L12
L45
FA linker
L12
A
L45
AB linker
β5
L45
β4
α1
L12
B
DNA(dT30)
180°
1
2
3
4
5
6
7
8
b.
c.
d.
e.
f.
g.
h.
i.
j.
k.
l.
m.
n.
o.
p.
q.
RPA
RPA+DNA
deuterium uptake
time (hr)
0.008 0.05 0.5 3 24
ns
*
**
***
13-25
Peptide 3
45-56
Peptide 4
67-73
Peptide 6
88-92
Peptide 7
212-218
Peptide 15
244-253
Peptide 16
246-256
Peptide 17
274-279
Peptide 19
277-303
Peptide 20
298-301
Peptide 21
297-307
Peptide 22
305-316
Peptide 25
349-357
Peptide 26
365-376
Peptide 28
395-415
Peptide 30
416-426
Peptide 32

Figure 2.3. Time-resolved HDX-MS analysis of DNA-induced conformational changes in the FAB half of RPA. a. Peptides identified in the MS analysis are mapped onto the crystal structure of PID^{70N}, DBD-A, and DBD-B. The F-AB linker and A-B linkers are denoted by dotted lines. DNA is depicted as black sticks. b-q. Deuterium exchange profiles of individual peptides from the FAB half of RPA. Error bars reflect standard deviations, calculated as described in Material and Methods. */**/*** are p-values calculated using a two-tailed unpaired t-test between the respective time points for the RPA and RPA+DNA samples. Each star denotes t-test values that differ by an order of magnitude. No significance indicated as 'ns'.

For DBD-A, six peptides were observed with varying HDX changes. Peptide15 is one of two loops (L12) at the cleft of the DNA binding pocket (**Figure 2.3f**). This region does not show rapid HDX but slower exposure to solvent is observed. The regions away from the cleft in the α2/β4 loop do not show changes in HDX (peptides 16, and 17; **Figure 2.3g, h**). Rapid DNA induced protection is observed behind the DNA binding cleft, and the tail-end of this region feeds into the linker that connects to DBD-B (peptides 19 and 20; **Figure 2.3i, j**). In the absence of DNA, peptides 21 and 22 from the AB-linker show slow and fast deuterium exchange overtime, respectively. Upon DNA binding an increase deuterium uptake is observed for both the peptides, however, the pattern was different (**Figures 2.3k & l**). The rate of exchange for peptide 21 was impacted at the first time point, whereas peptide 22 was impacted at later time points. This suggests that protection and dynamics are altered by DNA binding. Presumably, since the linker is located closer to the DNA binding cleft of DBD-B, greater deuterium uptake in these two peptides is likely directly tied to DNA binding (. Overlay of apo and DNA bound DBD-A show overall conformational changes upon DNA binding.

MS coverage for DBD-B yielded a total of 14 peptides (**Figure 2.1b**). The HDX changes in this domain were minimal upon DNA binding. Most peptides in this region

(peptides 23, 24, 27, 2918) do not show significant changes in HDX. Peptides 25 and 26 showed slower conformational changes upon DNA binding (Figure 3m & n). Peptide 26 in DBD-B (β3-strand) shows slow deuterium uptake kinetics with significantly decreased uptake (protection) at later time points with DNA. Regions within the DNA binding cleft (peptides 27, 28, 29, 30) have similar D_2O exchange patterns, with a minor decrease initially that is more pronounced at later time points upon DNA binding (**Figures 2.3o & p**). In the C-terminus of DBD-B (α2-helix; peptide 31), DNA binding leads to protection from D_2O exchange. Peptide 32 shows significant DNA-induced protection at the first time point (**Figure 2.3q**). Peptides corresponding to the BC-linker (peptides 32, 33 and 34) display dynamic changes with DNA dependent protection. Along with the DNA induced conformational changes, HDX kinetics in several regions of DBD-B fluctuate between time points, especially within the DNA binding cleft. Comparison of apo (PDB:1FGU) and DNA-bound (PDB:1JMC) DBD-B structures reveals several regions undergoing conformational changes. Specifically, largescale changes occur in the L12 and L45 loop regions.

In the context of the full-length RPA, our data support a more dynamic model for DBD-A and DBD-B engagement in agreement with recent single molecule fluorescence studies *(52)*. As expected, binding to DNA is observed as rapid HDX changes are observed for both DBDs, however, the extent of protection is not consistent with a stably bound configuration. DBD-A and DBD-B can adopt an array of configurations due to flexible and rotational effects of the AB linker *(36)*. Our HDX data is consistent with conformational flexibility of these domains. Residues 314 to 316 (peptides 24-25) and

395 to 401 (peptides 29-30) are at the interface between DBD-A and -B in the DNA bound structure and are solvent occluded. These residues show protection upon DNA binding (**Figures 2.2a, 2.3m & p**). Similarly, peptide 20 in DBD-A that includes the linker to DBD-B also supports similar DNA binding induced protection (**Figure 2.3j**). Thus, our observations agree with the existence of various configurational states of DBD-A and DBD-B *(36,63)*. There are limited contacts between DBD-A and -B with a flexible linker in apo structures which are brought together upon DNA binding. Time resolved deuterium exchange in the FAB half of RPA shows a gradual increase in deuterium uptake that is not linear. Furthermore, the amplitude of deuterium uptake does not reach saturation (**Figure 2.3**). Such deuterium exchange profiles are characteristic of protein-DNA interactions that are dynamic with rapid on and off rates *(59,64)*. The dynamic DNA binding properties of the FAB half are in agreement with recent single molecule studies that showcase multi-step dynamic interactions between DBD-A, DBD-B and ssDNA *(43,52)*.

Tri-C Shows Less Dynamic ssDNA-Dependent Changes in HDX

The trimerization core (Tri-C) is composed of DBD-C (RPA70), DBD-D (RPA32), PID^{32C} and DBD-E (**Figure 2.4a**). Structural and biochemical evidence suggests that Tri-C works as a single unit and binds to 24 nts *(30)*. DBD-C is connected to DBD-B by the BC-linker (13 residues) and can be envisioned as a hinge holding the two halves of the RPA together (**Figure 2.1a**). HDX analysis of DBD-C, the largest DNA binding domain in RPA, yields excellent coverage with total 24 peptides (**Figure 2.1b**). The HDX changes in DBD-C align well with expectations for a stable DNA binding

domain. Most differences in deuterium uptake upon DNA binding in DBD-C occurred on fast timescales and were of significant magnitude (**Figures 2.2a, 2.4**). Part of the BC-linker and the N-terminus of DBD-C (peptides 35, 36, 37, and 39) consistently show deuteration differences at every time point with a noticeable decrease in deuterium uptake upon DNA binding (**Figure 2.4b-e**). The N-terminus is less protected on DNA binding (peptide 38, Figure S5a). The BC-linker in the U. maydis RPA structure assumes two β-turn structures *(65)*. The first β-turn inserts into the DNA binding groove of DBD-B and is proposed to preclude DNA from adopting conformations observed in the 8 nt mode. In human RPA, this linker is shorter (13 residues) compared to U. maydis RPA (28 residues). However, the residues in the first β-turn are conserved (G435,435, V436,436, G437, G438). Our data likely reflects this structural transition from a flexible loop into a two β-turn upon binding to ssDNA. Furthermore, the loss in deuterium uptake amplitude upon DNA binding suggests occlusion of this region from solvent.

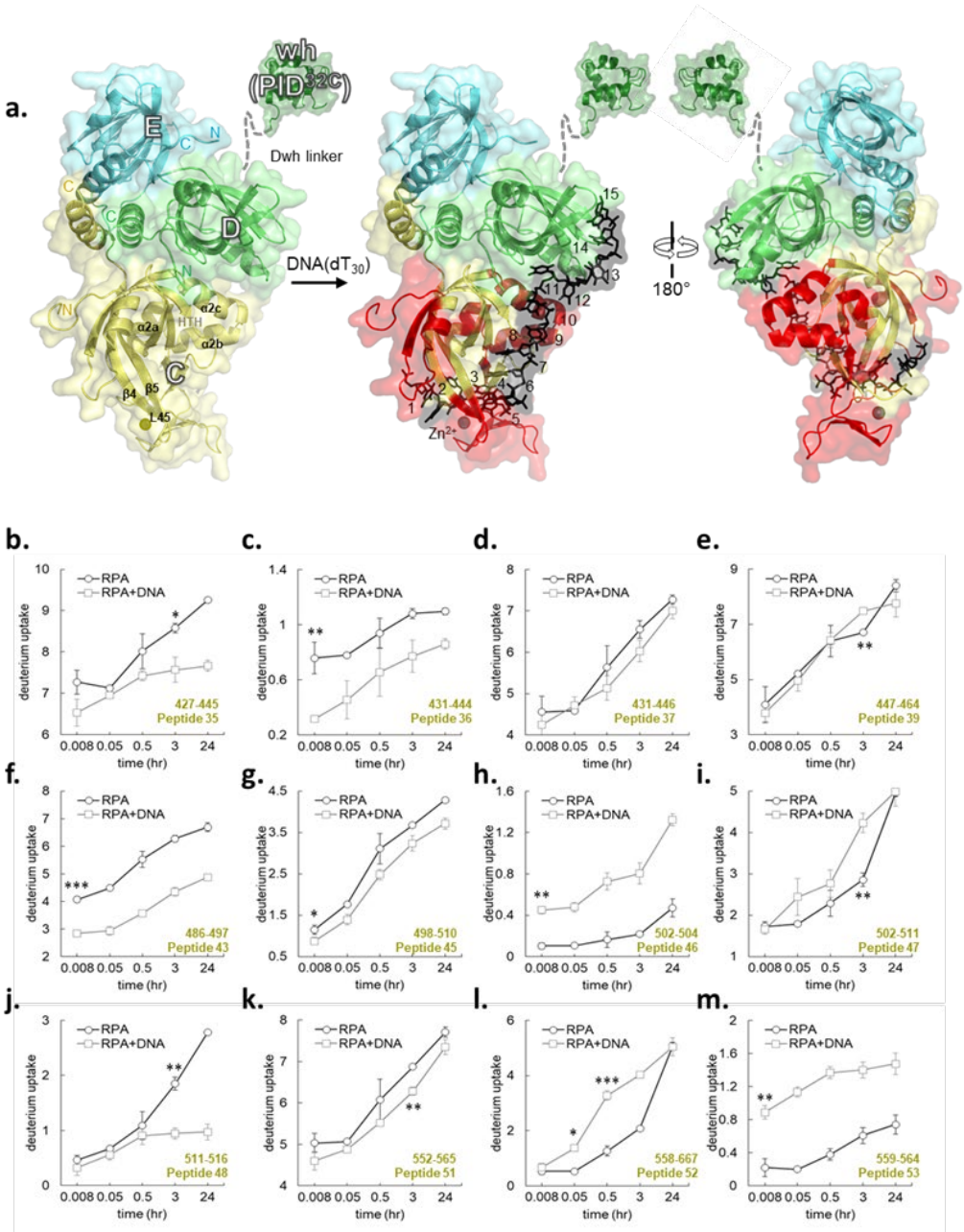

Figure 2.4. Time-resolved HDX-MS analysis of DNA-induced conformational changes in DBD-C.a. Crystal structure of the Tri-C core of human RPA composed of DBD-C, RPA32 and RPA14. The peptides identified in the MS analysis corresponding to DBD-C are denoted as in red. DNA is denoted as black sticks. b-m. Deuterium exchange profiles of individual peptides from DBD-D are shown. Error bars reflect standard deviations, calculated as described in Material and Methods. */**/*** are p-values calculated using a two-tailed unpaired t-test between the respective time points for the RPA and RPA+DNA

samples. Each star denotes t-test values that differ by an order of magnitude. p-values are denoted in Supplemental Table 2. No significance indicated as 'ns'.

Upon DNA binding, major changes in the DBD-C structure are also observed around the Zn^{2+}-finger motif. Four cysteines (C481, C486, C500, and C503) that coordinate zinc binding match peptides 41-43 and 45-46 (**Figure 2.4f-h**). Correspondingly, in our HDX analysis, peptides 43 to 48 have major changes in deuterium uptake (**Figure 2.4f-j**). The inner surface of DBD-C containing the DNA binding cleft shows significant decrease in deuterium exchange levels suggesting protection upon DNA binding and displays rapid exchange kinetics (peptides 43, 45; **Figures 2.4f-g**). The decrease in D_2O uptake reflects occlusion of this region from solvent upon DNA binding and the rapid kinetics denote direct DNA induced changes. Similarly, the solvent exposed outer surface region of the Zn^{2+}-finger region showed a robust increase in deuterium uptake upon DNA binding with fast exchange kinetics (peptides 46, 47; **Figures 2.4h-i**).

Significant differences in HDX was also observed in the helix-turn-helix motif (Supplemental Figure S6, HTH; α2a- α2b- α2c; 533 to 566 aa), a region underneath the DNA binding cleft of DBD-C (peptides 49 to 54, **Figures 2.4k-l**). Peptide 48 is protected in presence of DNA, becoming more dynamic (exchange on a long timescale), having slow movements when DNA is absent (**Figure 2.4j**). Upon DNA binding, the N-terminal half of HTH motif (peptides 49-51) showed reduction in deuterium uptake levels. The first incubation time point that shows statistical difference in deuterium uptake from the RPA alone are indicated for peptides 49-51 respectively (**Figure 2.4k**). In contrast, the C-terminal half of HTH motif (peptides 52 to 54) displayed increased deuterium uptake and

faster exchange kinetics (**Figures 2.4l-m**). The first time point post DNA addition that shows statistical difference in deuterium uptake are indicated for peptides 52-54 respectively (**Figures 2.4l-m**). Simultaneous protection and exposure in D_2O exchange within this region suggests correlated conformational movements. That is, occlusion from solvent on the N-terminal side and exposure to solvent on the C-terminal side of HTH motif. It can be envisioned that DNA binding occludes the N-terminal end from deuteration by compaction which leads to conformational change that extends the C-terminal half. Indeed, in our modelled structure, we do see an extended conformation of HTH C-terminal half which can now contact DBD-D and DNA through van der Waals interactions *(65)*. A similar interaction between DBD-C and DBD-D is observed in the cryoEM structure of S. cerevisiae RPA *(53)*. The L45 loop, connecting β-strand 4 and β-strand 5, is a flexible region in DBD-C and harbors Y581, a conserved residue that base-stacks with DNA. Another conserved aromatic residue F532 resides in β3-strand of DBD-C. Other residues in this region (R575, K577, Y581, and R586) also make extensive contacts with ssDNA. Electron density for L45 in the structure of U. maydis Tri-C is poorly defined and suggests a dynamic region that is likely ordered upon binding to DNA. Peptide 55 corresponds to the L45 loop and is protected on DNA binding at early time points with a steady increase in HDX. Our data for DBD-C showcase robust and stable changes in deuteration levels upon interaction with DNA. In comparison to the highly dynamic FAB-half of RPA, DBD-C appears to stably interact with DNA with fewer dynamic changes.

RPA32 harbors a disordered N-terminal region which can be heavily phosphorylated, DBD-D, and PID^{32C}. 17 total peptides for RPA32 were identified and analyzed: Peptides 1 to 5 cover the N-terminal region and show small changes in HDX upon DNA binding. Peptides 1 and 5 show rapid DNA binding dependent changes whereas peptides 2 and 4 do not (**Figure 2.5b-d**). These data suggest that this region is undergoing conformational changes upon DNA binding, but it is interesting to note that this region is not part of the DNA binding domain in the structures. Peptides 9 to 13 encompass DBD-D and correspondingly show fast exchange kinetics at early time points, indicative of direct DNA binding induced changes in the structure (**Figure 2.5f-j**). The data are consistent with DNA dependent occlusion/compaction of the DNA binding site. One key structural motif in DBD-D is the β4-β5 hairpin (i.e. β4-L45-β5) which is an extension of the DNA binding site and displays rapid exchange (peptide 9 and 10) with a two-phase deuterium exchange profile observed for the β5 peptide (peptide 10). Upon DNA binding the C-terminal α-helix shows exposed (peptide 11) and protected (peptide 12) profiles (Figures 5h-i). These changes suggest that the α-helical bundle of Tri-C is dynamic and propagates DNA-induced conformational changes.

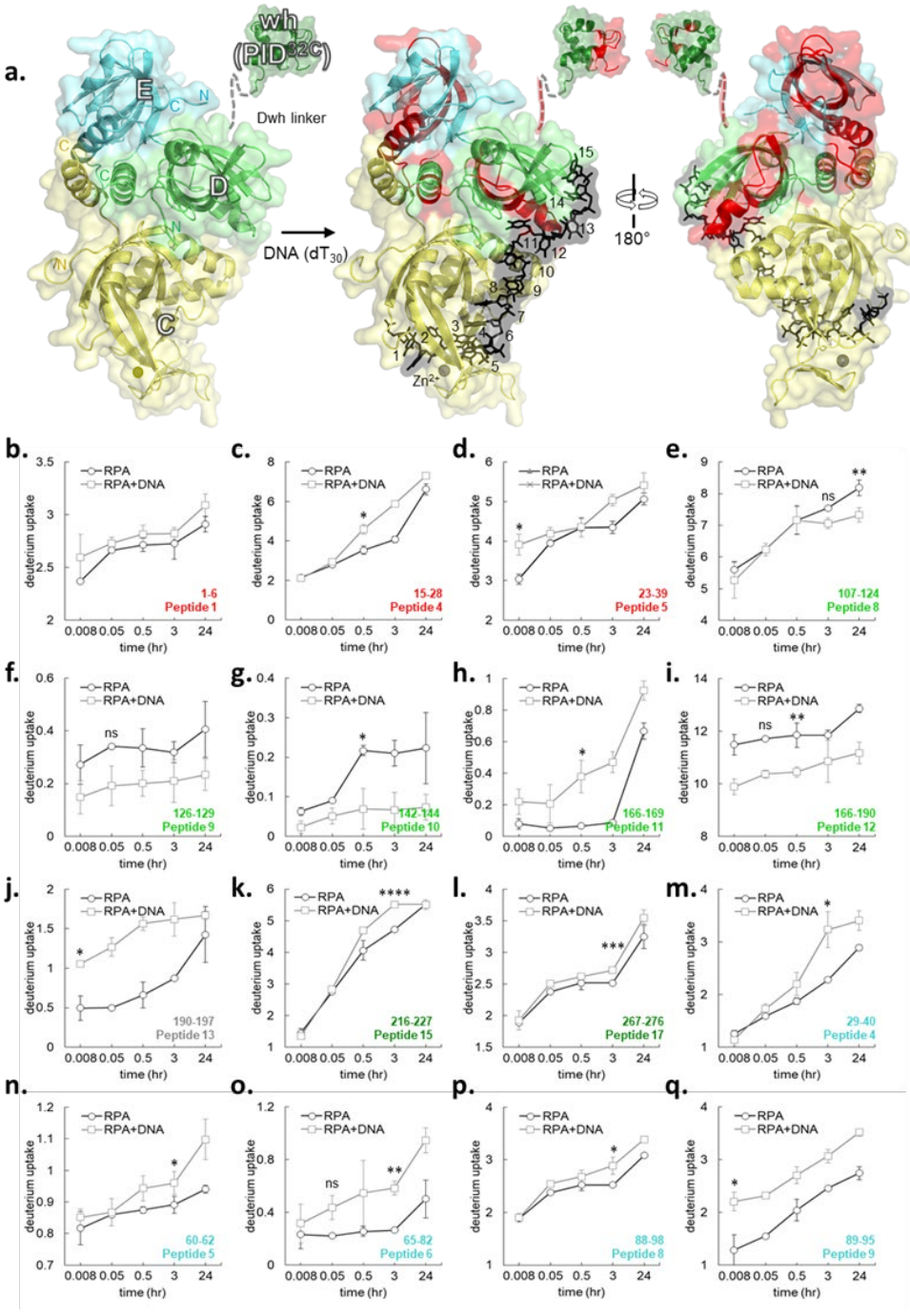

Figure 2.5. Time-resolved HDX-MS analysis of DNA-induced conformational changes in RPA32 and RPA14. a. Crystal structure of the Tri-C core of human RPA composed of DBD-C, RPA32 and RPA14. The peptides identified in the MS analysis corresponding to RPA32 and RPA14 are denoted in red. DNA is depicted as black sticks. b-q. Deuterium

exchange profiles of individual peptides from RPA32 and RPA14 are shown. Error bars reflect standard deviations, calculated as described in Material and Methods. */**/*** are p-values calculated using a two-tailed unpaired t-test between the respective time points for the RPA and RPA+DNA samples. Each star denotes t-test values that differ by an order of magnitude. p-values are denoted in Supplemental Table 2. No significance indicated as 'ns'.

Interestingly, PID^{32C} (peptide 15, 16, 17) and its linker region (peptide 13) also show changes in deuteration in the presence of DNA (**Figures 2.5j-l**). Peptide 13 corresponds to the linker connecting DBD-D and PID^{32C} (**Figures 2.2b and 2.5a**) and shows a significant increase in deuterium uptake suggesting solvent exposure (**Figure 2.5j**). One possibility is that interactions exist between PID^{32C} and DBD-D are these interactions are perturbed upon DNA binding. Altered uptake patterns are also observed in the tail-end of the helix in DBD-D (described above) and the linker. The deuteration profile of the PID^{32C} peptides do not show robust changes (peptide 15-17, **Figures 2.5k-l**) similar to the observations for PID^{70N} (**Figures 2.3b-e**). An overlay of apo (PDB:1L10) and DNA bound (PDB:4GNX) DBD-D structures reveals mild changes. Thus, we conclude that both the protein interaction domains in human RPA do not interact directly with ssDNA.

DBD-E is the third component of Tri-C. In the crystal structures, no direct interactions are observed between DNA and DBD-E. However, crosslinking experiments have suggested DNA interactions in this domain *(50)*. Interestingly, in the cryoEM structure of multiple yeast RPA molecules on ssDNA, DBD-A of one RPA molecule interacts with DBD-E of the neighboring RPA *(53)*. This interaction is further stabilized upon phosphorylation at position S178 in the FA-linker closer to DBD-A. Thus, DBD-E is an integral component of cooperative binding of RPA molecules upon phosphorylation.

Eleven peptides were tracked in DBD-E. Our data show both fast (peptides 5-6, 8, 10) and slow (peptides 3-4, 7, 9, 10-11) D_2O exchange for DBD-E. Several of the corresponding peptides (4-6, and 8-9) show DNA dependent solvent exposure (**Figures 2.5m-q**). Peptides 3 and 4 are in the β1-strand region, and significant solvent exposure was observed for peptide 4 (**Figure 2.5m**). Peptide 5 is part of L34 loop in DBD-E and also displays significant solvent exposure kinetics at 3h (**Figure 2.5n**). There are two overlapping peptides for the β4-β5 hairpin; where peptide 6 shows rapid and robust deuterium uptake while significant change was observed for peptide 7 (**Figure 2.5o**). This suggests solvent exposure of the β4 and L45 regions in DBD-E. Analogous regions in other DBDs (DBD-B peptides 18-19; DBD-C peptides 41-44; DBD-D peptides 9-10) show large conformational changes upon DNA binding. The loop between β5 and α2 region in DBD-E was covered by peptides 8, 9 and 10. Peptide 8 that overlaps both 9 and 10 showed significant and delayed deuterium uptake (**Figure 2.5p**). Interestingly, peptides 9 and 10 showed greater exchange reflecting exposure and protection respectively (**Figure 2.5q**). Thus, the entire DBD-E is exposed and the beginning of α2 (peptide 10) is protected (Figure 2b and Supplemental Figure S8c). However, most of the α2-helical region in DBD-E (peptide 11) is dynamic without any significant change in deuteration on DNA binding. The dynamic characteristics of the C-terminal α-helix in DBD-E is similar to the C-terminal α-helix in DBD-D (**Figures 2.5h-i**). Both these α-helices are part of the α-helix bundle in Tri-C.

Taken together, our data show that the Tri-C region of RPA is more stable compared to the FAB region. The Tri-C region of RPA shows rapid deuterium uptake in

the presence of DNA and the deuteration levels suggest that the structures (H-bonding networks) are stable compared to the FAB region. Much of this stability arises from the interaction of DBD-C and DNA, with dynamic changes observed in DBD-D and DBD-E. Thus, within Tri-C, DBD-D and DBD-E are relatively dynamic in comparison to DBD-C but are less dynamic compared to DBD-A and -B. Our data support a model for RPA where DBDs-A and -B form the 'dynamic half' and DBDs-C, -D and -E form a 'less dynamic half' of RPA. These conclusions do not suggest a loss in DNA binding for the FAB half but likely reflect faster intrinsic on-off rates for ssDNA interactions.

Discussion

One of the primary roles of RPA is to protect ssDNA from nucleases and act as a platform to recruit other proteins. This is an inherently dynamic process in which RPA hands off the DNA to the incoming enzyme. The flexibility in RPA structure arises from the compartmentalization of its DBDs and PIDs which are tethered together by flexible linkers. The modular architecture allows the domains of RPA to employ both independent and coordinated DNA interactions that are tailored for specific functional outcomes *(28,29,43)*.

Canonical models for RPA-DNA interactions primarily arose from biochemical investigations of isolated or select pairs of DBDs. DBDs-A and -B are defined as high affinity DNA binding domains. Tri-C is considered the weaker DNA binding portion of RPA. Recent single molecule and structural data suggest that these high-affinity DBDs might be more dynamic. DBD-A and DBD-B dissociate from DNA and interact with DBD-E from the neighboring molecule when multiple RPA molecules are visualized

using cryoEM *(53)*. The Tri-C is more ordered in these structures as well. In single molecule experiments where the DNA binding dynamics of individual DBDs are monitored, DBD-A and DBD-D can form four distinct configurations on DNA and transition between these states. In the presence of an RPA interacting protein, one of these states is not formed *(43,52)*. These studies provide direct support for the dynamics-based model wherein the dynamics of individual DBDs provides a mechanism to access RPA-bound ssDNA whilst not completely dissociating RPA *(28,29,37)*. The single molecule experiments do not provide such information for the other DBDs or PIDs (due to technical limitations). Thus, a complete picture of the dynamics of all the domains of RPA and how they are rearranged in the presence of DNA is lacking.

The HDX data presented here provide a coarse-grain structural and kinetic landscape of the DNA-driven conformational changes (**Figure 2.6a**). The HDX profiles of all five DBDs (A, B, C, D, E) have altered exchange kinetics on a fast timescale in the presence of DNA. This result is consistent with rapid on-rates for the individual DBDs binding to DNA. However, the exchange patterns of the HDX signal varies for the individual DBDs. The HDX of DBDs-A and -B rapidly re-equilibrate suggesting rapid on-off kinetics. Thus, these domains and their interactions are more dynamic in nature as denoted by the early timepoints in the HDX analysis. DBDs-C, -D and -E also show large scale deuterium uptake profiles and show slow re-equilibration. HDX data at the later timepoints reveal that deuterium uptake levels off over a period of 24 hours for many of their peptides suggesting formation of stable H-bonding networks (**Figures 2.4 and 2.5**).

Thus, the DBDs in the Tri-C region are less dynamic compared to the ones in the FAB half of RPA.

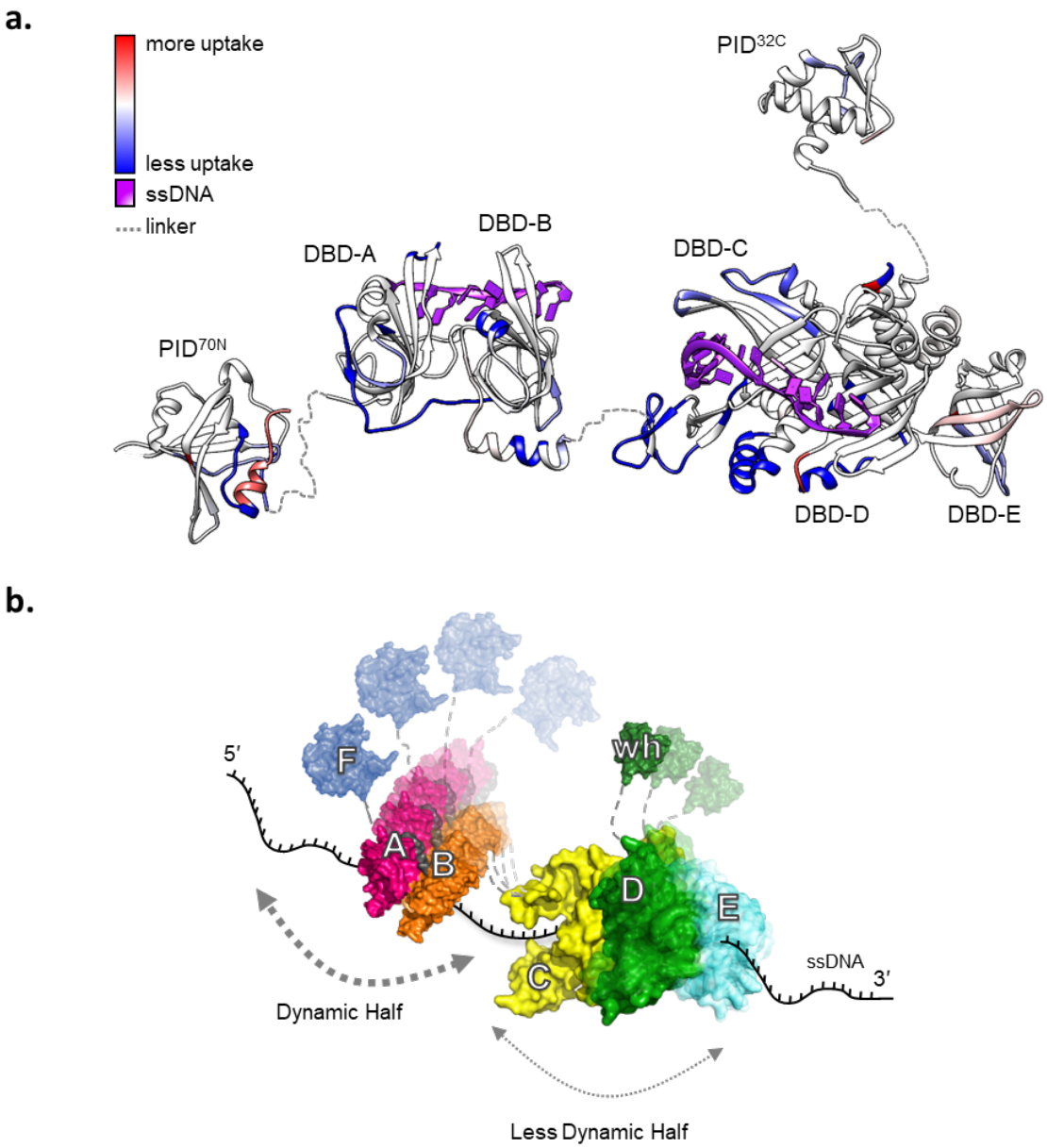

Figure 2.6. RPA binds to ssDNA using dynamic and less dynamic halves. a. Global HDX changes in human RPA upon binding to DNA are mapped onto a composite structure. Blue and red denote the difference in scale of deuterium uptake for the unbound (RPA) versus DNA-bound (RPA+DNA) forms of RPA. DNA binding primarily drives protection of regions from deuterium uptake as they become occluded from solvent. b. An updated dynamics-based model for RPA+DNA interactions is depicted. RPA is

proposed to function as two halves (FAB and CDE/Tri-C). PID^{70N} (F domain), DBD-A and DBD-B form the dynamic half and have short residence on time. DBD-C, RPA32 (DBD-D) and RPA14 (DBD-E) form the less dynamic half and have longer residence time on DNA owing to the stability of DNA binding arising from the larger DBD-C domain. The refined model allows RPA to be stably bound to DNA using all DBDs while providing access to both ends of the DNA due to the dynamics of the individual DBDs. PID^{70N} and PID^{32C} primarily serve to bind and recruit RPA-interacting proteins onto DNA.

We do not observe fast exchange kinetics or robust deuterium changes in either PID^{70N} or PID^{32C} suggesting that the primary roles of these domains are to interact with RIPs. Binding of p53 to the cleft in PID^{70N} has been shown to be outcompeted by DNA, and thus proposed to possess very weak DNA binding properties *(51)*. But more importantly, PID^{70N} serves as a high affinity binding site for more than two dozen proteins including p53 *(51)* and Dna2 *(66)*. In addition, PID^{70N} interacts with the N-terminal phosphorylation motif of RPA32 *(67)*. Thus, the HDX changes we observe in the later timescales likely originate from the release of such intra-RPA:PID^{70N} interactions upon DNA binding to DBD-C and DBD-D, and not from direct DNA binding to PID^{70N}.

Biophysical characterization of full-length RPA suggests that DNA binding is highly dynamic and that the interaction of individual DBDs is not sequential as previously envisioned *(46)*. The correlated movements/dynamics of the DBDs (and likely PIDs) enable rapid RPA diffusion on ssDNA (~5,000 $nt^2\ sec^{-1}$ at 37 °C) and efficiently melt small hairpins and duplex regions as it diffuses *(44)*. RPA–DNA complexes have also been shown to exchange rapidly in the presence of free RPA *(68,69)*. Transition between specific DNA binding modes are thought to promote such mechanical movements of RPA. An initial 8-10 nt is formed first followed by transition to a 20 nt

mode and finally resulting in the 30 nt mode where all DBDs are DNA-bound *(70)*. RPA dynamically forms at least two kinetically distinct complexes on DNA with different stabilities that are microscopically dissociating and re-associating *(71)*.

Our data support a DNA binding dynamics-based model for RPA-DNA interactions. We envision RPA to function as two dynamic halves partitioned as the FAB half and the Tri-C half (**Figure 2.6b**). The FAB half is dynamic with rapid on-off kinetics on ssDNA and the Tri-C half is less-dynamic with more stable interactions with DNA, primarily driven by the stability of DBD-C. These data are not to interpreted as the FAB half having poor DNA binding activity as the domains likely maintain high affinity interactions whilst possessing rapid on-off kinetics. These findings do suggest that the DNA buried under the FAB half of RPA should be transiently accessible to RIPs.

The HDX data do not support a role for the PIDs in mediating direct DNA interactions, but intra-domain interactions exist between the PIDs and the DBDs. These interactions are released as conformational changes are driven upon DNA binding as slow exchange HDX is observed for both PIDs. Such changes would allow RPA to maintain high affinity interactions through stable binding by DBD-C and contributions from other parts of Tri-C. The dynamic interactions of DBDs-A and -B would enable RIPs that interact with PID^{70N} to gain access to the 5′ end of the RPA-coated DNA. Similarly, even the less-dynamic DBDs-D and -E can be remodeled by RIPs that interact with PID^{32C} and bind to the 3′ of the RPA-coated DNA as previously observed for Rad52 *(43,52)*. When multiple RPA molecules are bound on the DNA, the dynamic nature of the DBDs would enable RIPs to bind between multiple RPA molecules bound to the DNA.

References

(23) Kim, C.; F. Paulus, B.; S. Wold, M. Interactions of Human Replication Protein A with Oligonucleotides. *Biochemistry* **2002**, *33* (47), 14197–14206. https://doi.org/10.1021/bi00251a031.

(24) Namiki, Y.; Zou, L. ATRIP Associates with Replication Protein A-Coated SsDNA through Multiple Interactions. *Proc Natl Acad Sci PNAS* **2006**, *103*, 580–585.

(25) Zou, L.; Elledge, S. J. Sensing DNA Damage through ATRIP Recognition of RPA-SsDNA Complexes. *Science (80-.).* **2003**, *300*, 1542–1548.

(26) Maréchal, A.; Zou, L. RPA-Coated Single-Stranded DNA as a Platform for Post-Translational Modifications in the DNA Damage Response. *Cell Res.* **2015**, *25* (1), 9–23. https://doi.org/10.1038/cr.2014.147.

(27) 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.; Chazin, W. J. A Key Interaction with RPA Orients XPA in NER Complexes. *Nucleic Acids Res.* **2020**, *48* (4), 2173–2188. https://doi.org/10.1093/nar/gkz1231.

(28) Fanning, E.; Klimovich, V.; Nager, A. R. A Dynamic Model for Replication Protein A (RPA) Function in DNA Processing Pathways. *Nucleic Acids Res.* **2006**, *34* (15), 4126–4137. https://doi.org/10.1093/nar/gkl550.

(29) Sugitani, N.; Chazin, W. J. Characteristics and Concepts of Dynamic Hub Proteins in DNA Processing Machinery from Studies of RPA. *Prog. Biophys. Mol. Biol.* **2015**, *117* (2–3), 206—211. https://doi.org/10.1016/j.pbiomolbio.2014.12.001.

(30) Bochkareva, E.; Korolev, S.; Bochkarev, A. The Role for Zinc in Replication Protein A. *J. Biol. Chem.* **2000**, *275* (35), 27332–27338. https://doi.org/10.1074/jbc.M000620200.

(31) Bochkareva, E.; Belegu, V.; Korolev, S.; Bochkarev, A. Structure of the Major Single-Stranded DNA-Binding Domain of Replication Protein A Suggests a Dynamic Mechanism for DNA Binding. *EMBO J.* **2001**, *20* (3), 612–618. https://doi.org/10.1093/emboj/20.3.612.

(32) Bochkareva, E.; Korolev, S.; Lees-Miller, S. P.; Bochkarev, A. Structure of the RPA Trimerization Core and Its Role in the Multistep DNA-Binding Mechanism of RPA. *EMBO J.* **2002**, *21* (7), 1855–1863. https://doi.org/10.1093/emboj/21.7.1855.

(33) Nakaya, R.; Takaya, J.; Onuki, T.; Moritani, M.; Nozaki, N.; Ishimi, Y. Identification of Proteins That May Directly Interact with Human RPA. *J.*

Biochem. **2010**, *148* (5), 539—547. https://doi.org/10.1093/jb/mvq085.

(34) Dueva, R.; Iliakis, G. Replication Protein A: A Multifunctional Protein with Roles in DNA Replication, Repair and Beyond. *NAR Cancer* **2020**, *2* (3). https://doi.org/10.1093/narcan/zcaa022.

(35) Mer, G.; Bochkarev, A.; Gupta, R.; Bochkareva, E.; Frappier, L.; Ingles, C. J.; Edwards, A. M.; Chazin, W. J. Structural Basis for the Recognition of DNA Repair Proteins UNG2, XPA, and RAD52 by Replication Factor RPA. *Cell* **2000**, *103* (3), 449–456. https://doi.org/10.1016/s0092-8674(00)00136-7.

(36) Brosey, C. A.; Soss, S. E.; Brooks, S.; Yan, C.; Ivanov, I.; Dorai, K.; Chazin, W. J. Functional Dynamics in Replication Protein A DNA Binding and Protein Recruitment Domains. *Structure* **2015**, *23* (6), 1028–1038. https://doi.org/10.1016/j.str.2015.04.008.

(37) Brosey, C. A.; Yan, C.; Tsutakawa, S. E.; Heller, W. T.; Rambo, R. P.; Tainer, J. A.; Ivanov, I.; Chazin, W. J. A New Structural Framework for Integrating Replication Protein A into DNA Processing Machinery. *Nucleic Acids Res.* **2013**, *41* (4), 2313–2327. https://doi.org/10.1093/nar/gks1332.

(38) Iftode, C.; Borowiec, J. A. 5' --> 3' Molecular Polarity of Human Replication Protein A (HRPA) Binding to Pseudo-Origin DNA Substrates. *Biochemistry* **2000**, *39* (39), 11970–11981. https://doi.org/10.1021/bi0005761.

(39) de Laat, W. L.; Appeldoorn, E.; Sugasawa, K.; Weterings, E.; Jaspers, N. G.; Hoeijmakers, J. H. DNA-Binding Polarity of Human Replication Protein A Positions Nucleases in Nucleotide Excision Repair. *Genes Dev.* **1998**, *12* (16), 2598–2609. https://doi.org/10.1101/gad.12.16.2598.

(40) Bae, S. H.; Bae, K. H.; Kim, J. A.; Seo, Y. S. RPA Governs Endonuclease Switching during Processing of Okazaki Fragments in Eukaryotes. *Nature* **2001**, *412* (6845), 456–461. https://doi.org/10.1038/35086609.

(41) Sugawara, N.; Haber, J. E. Characterization of Double-Strand Break-Induced Recombination: Homology Requirements and Single-Stranded DNA Formation. *Mol. Cell. Biol.* **1992**, *12* (2), 563–575. https://doi.org/10.1128/mcb.12.2.563.

(42) Bastin-Shanower, S. A.; Brill, S. J. Functional Analysis of the Four DNA Binding Domains of Replication Protein A. The Role of RPA2 in SsDNA Binding. *J. Biol. Chem.* **2001**, *276* (39), 36446–36453. https://doi.org/10.1074/jbc.M104386200.

(43) Caldwell, C. C.; Spies, M. Dynamic Elements of Replication Protein A at the Crossroads of DNA Replication, Recombination, and Repair. *Crit. Rev. Biochem. Mol. Biol.* **2020**, *55* (5), 482–507. https://doi.org/10.1080/10409238.2020.1813070.

(44) Nguyen, B.; Sokoloski, J.; Galletto, R.; Elson, E. L.; Wold, M. S.; Lohman, T. M. Diffusion of Human Replication Protein A along Single-Stranded DNA. *J. Mol. Biol.* **2014**, *426* (19), 3246–3261. https://doi.org/10.1016/j.jmb.2014.07.014.

(45) Antony, E.; Lohman, T. M. Dynamics of E. Coli Single Stranded DNA Binding (SSB) Protein-DNA Complexes. *Semin. Cell Dev. Biol.* **2019**, *86*, 102–111. https://doi.org/10.1016/j.semcdb.2018.03.017.

(46) Cai, L.; Roginskaya, M.; Qu, Y.; Yang, Z.; Xu, Y.; Zou, Y. Structural Characterization of Human RPA Sequential Binding to Single-Stranded DNA Using SsDNA as a Molecular Ruler. *Biochemistry* **2007**, *46* (28), 8226–8233. https://doi.org/10.1021/bi7004976.

(47) Walther, A. P.; Gomes, X. V; Lao, Y.; Lee, C. G.; Wold, M. S. Replication Protein A Interactions with DNA. 1. Functions of the DNA-Binding and Zinc-Finger Domains of the 70-KDa Subunit. *Biochemistry* **1999**, *38* (13), 3963–3973. https://doi.org/10.1021/bi982370u.

(48) Wyka, I. M.; Dhar, K.; Binz, S. K.; Wold, M. S. Replication Protein A Interactions with DNA: Differential Binding of the Core Domains and Analysis of the DNA Interaction Surface. *Biochemistry* **2003**, *42* (44), 12909–12918. https://doi.org/10.1021/bi034930h.

(49) Arunkumar, A. I.; Stauffer, M. E.; Bochkareva, E.; Bochkarev, A.; Chazin, W. J. Independent and Coordinated Functions of Replication Protein A Tandem High Affinity Single-Stranded DNA Binding Domains. *J. Biol. Chem.* **2003**, *278* (42), 41077–41082. https://doi.org/10.1074/jbc.M305871200.

(50) Salas, T. R.; Petruseva, I.; Lavrik, O.; Saintomé, C. Evidence for Direct Contact between the RPA3 Subunit of the Human Replication Protein A and Single-Stranded DNA. *Nucleic Acids Res.* **2009**, *37* (1), 38–46. https://doi.org/10.1093/nar/gkn895.

(51) Bochkareva, E.; Kaustov, L.; Ayed, A.; Yi, G.-S.; Lu, Y.; Pineda-Lucena, A.; Liao, J. C. C.; Okorokov, A. L.; Milner, J.; Arrowsmith, C. H.; Bochkarev, A. Single-Stranded DNA Mimicry in the P53 Transactivation Domain Interaction with Replication Protein A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15412—15417. https://doi.org/10.1073/pnas.0504614102.

(52) Pokhrel, N.; Caldwell, C. C.; Corless, E. I.; Tillison, E. A.; Tibbs, J.; Jocic, N.; Tabei, S. M. A.; Wold, M. S.; Spies, M.; Antony, E. Dynamics and Selective Remodeling of the DNA-Binding Domains of RPA. *Nat. Struct. Mol. Biol.* **2019**, *26* (2), 129–136. https://doi.org/10.1038/s41594-018-0181-y.

(53) Yates, L. A.; Aramayo, R. J.; Pokhrel, N.; Caldwell, C. C.; Kaplan, J. A.; Perera, R. L.; Spies, M.; Antony, E.; Zhang, X. A Structural and Dynamic Model for the

Assembly of Replication Protein A on Single-Stranded DNA. *Nat. Commun.* **2018**, *9* (1), 5447. https://doi.org/10.1038/s41467-018-07883-7.

(54) van Erp, P. B. G.; Patterson, A.; KANT, R.; Berry, L.; Golden, S. M.; Forsman, B. L.; Carter, J.; Jackson, R. N.; Bothner, B.; Wiedenheft, B. Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guide Cascade Complex. *ACS Chem. Biol.* **2017**, acschembio.7b00649. https://doi.org/10.1021/acschembio.7b00649.

(55) Berry, L.; Poudel, S.; Tokmina-Lukaszewska, M.; Colman, D. R.; Nguyen, D. M. N.; Schut, G. J.; Adams, M. W. W.; Peters, J. W.; Boyd, E. S.; Bothner, B. H/D Exchange Mass Spectrometry and Statistical Coupling Analysis Reveal a Role for Allostery in a Ferredoxin-Dependent Bifurcating Transhydrogenase Catalytic Cycle. *Biochim. Biophys. Acta - Gen. Subj.* **2018**, *1862* (1), 9–17. https://doi.org/10.1016/j.bbagen.2017.10.002.

(56) Knox, R.; Lento, C.; Wilson, D. J. Mapping Conformational Dynamics to Individual Steps in the TEM-1 β-Lactamase Catalytic Mechanism. *J. Mol. Biol.* **2018**, *430* (18 Pt B), 3311–3322. https://doi.org/10.1016/j.jmb.2018.06.045.

(57) Zheng, J.; Chang, M. R.; Stites, R. E.; Wang, Y.; Bruning, J. B.; Pascal, B. D.; Novick, S. J.; Garcia-Ordonez, R. D.; Stayrook, K. R.; Chalmers, M. J.; Dodge, J. A.; Griffin, P. R. HDX Reveals the Conformational Dynamics of DNA Sequence Specific VDR Co-Activator Interactions. *Nat. Commun.* **2017**, *8* (1), 923. https://doi.org/10.1038/s41467-017-00978-7.

(58) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* **2019**, *16* (7), 595–602. https://doi.org/10.1038/s41592-019-0459-y.

(59) Graham, B. W.; Tao, Y.; Dodge, K. L.; Thaxton, C. T.; Olaso, D.; Young, N. L.; Marshall, A. G.; Trakselis, M. A. DNA Interactions Probed by Hydrogen-Deuterium Exchange (HDX) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Confirm External Binding Sites on the Minichromosomal Maintenance (MCM) Helicase. *J. Biol. Chem.* **2016**, *291* (24), 12467–12480. https://doi.org/10.1074/jbc.M116.719591.

(60) Zhang, J.; Li, J.; Craig, T. A.; Kumar, R.; Gross, M. L. Hydrogen-Deuterium

Exchange Mass Spectrometry Reveals Calcium Binding Properties and Allosteric Regulation of Downstream Regulatory Element Antagonist Modulator (DREAM). *Biochemistry* **2017**, *56* (28), 3523—3530. https://doi.org/10.1021/acs.biochem.7b00100.

(61) Bochkarev, A.; Pfuetzner, R. A.; Edwards, A. M.; Frappier, L. Structure of the Single-Stranded-DNA-Binding Domain of Replication Protein A Bound to DNA. *Nature* **1997**, *385* (6612), 176–181. https://doi.org/10.1038/385176a0.

(62) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

(63) Pretto, D. I.; Tsutakawa, S.; Brosey, C. A.; Castillo, A.; Chagot, M.-E.; Smith, J. A.; Tainer, J. A.; Chazin, W. J. Structural Dynamics and Single-Stranded DNA Binding Activity of the Three N-Terminal Domains of the Large Subunit of Replication Protein A from Small Angle X-Ray Scattering. *Biochemistry* **2010**, *49* (13), 2880—2889. https://doi.org/10.1021/bi9019934.

(64) Corrales-Guerrero, L.; He, B.; Refes, Y.; Panis, G.; Bange, G.; Viollier, P. H.; Steinchen, W.; Thanbichler, M. Molecular Architecture of the DNA-Binding Sites of the P-Loop ATPases MipZ and ParA from Caulobacter Crescentus. *Nucleic Acids Res.* **2020**, *48* (9), 4769–4779. https://doi.org/10.1093/nar/gkaa192.

(65) Fan, J.; Pavletich, N. P. Structure and Conformational Change of a Replication Protein A Heterotrimer Bound to SsDNA. *Genes Dev.* **2012**, *26* (20), 2337–2347.

(66) Zhou, C.; Pourmal, S.; Pavletich, N. P. Dna2 Nuclease-Helicase Structure, Mechanism and Regulation by Rpa. *Elife* **2015**, *4*. https://doi.org/10.7554/eLife.09832.

(67) Binz, S. K.; Lao, Y.; Lowry, D. F.; Wold, M. S. The Phosphorylation Domain of the 32-KDa Subunit of Replication Protein A (RPA) Modulates RPA-DNA Interactions. Evidence for an Intersubunit Interaction. *J. Biol. Chem.* **2003**, *278* (37), 35584–35591. https://doi.org/10.1074/jbc.M305388200.

(68) Pokhrel, N.; Origanti, S.; Davenport, E. P.; Gandhi, D.; Kaniecki, K.; Mehl, R. A.; Greene, E. C.; Dockendorff, C.; Antony, E. Monitoring Replication Protein A (RPA) Dynamics in Homologous Recombination through Site-Specific Incorporation of Non-Canonical Amino Acids. *Nucleic Acids Res.* **2017**, *45* (16), 9413–9426. https://doi.org/10.1093/nar/gkx598.

(69) Gibb, B.; Ye, L. F.; Gergoudis, S. C.; Kwon, Y.; Niu, H.; Sung, P.; Greene, E. C. Concentration-Dependent Exchange of Replication Protein A on Single-Stranded

DNA Revealed by Single-Molecule Imaging. *PLoS One* **2014**, *9* (2), e87922.

(70) Kumaran, S.; Kozlov, A. G.; Lohman, T. M. Saccharomyces Cerevisiae Replication Protein A Binds to Single-Stranded DNA in Multiple Salt-Dependent Modes. *Biochemistry* **2006**, *45* (39), 11958–11973. https://doi.org/10.1021/bi060994r.

CHAPTER THREE

DYNAMICS OF HEPATITIS B VIRUS CAPSID PROTEIN DIMER REGULATES ASSEMBLY THROUGH AN ALLOSTERIC NETWORK

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Angela Patterson

Contributions: Planned and performed HDX-MS and NMS experiments, analyzed data, and wrote the manuscript.

Co-Author: Zhongchao Zhao

Contributions: Purified the proteins studied in the manuscript, as well as, revising the manuscript.

Co-Author: Elizabeth Waymire

Contributions: Performed DSF experiment, and contributed to the writing and revision of the manuscript.

Co-Author: Adam Zlotnick

Contributions: Supervised the purification of the proteins used in the manuscript, and revised the manuscript.

Co-Author: Brian Bothner

Contributions: Supervised the HDX-MS, NMS, and DSF experiments, contributed to the writing and revision of the manuscript.

Manuscript Information

Angela Patterson, Zhongchao Zhao, Elizabeth Waymire, Adam Zlotnick, and Brian Bothner

ACS Chemical Biology

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__x_ Published in a peer-reviewed journal

ACS Chemical Biology
Volume 15, Issue 8
DOI: 10.1021/acschembio.0c00481

Abstract

While there is an effective vaccine for Human Hepatitis B Virus (HBV), 257 million people have chronic infections for which there is no cure. The assembly process for the viral capsid is a potential therapeutic target. In order to understand the capsid assembly process, we investigated the dimeric building blocks of the capsid. To understand what blocks assembly, we took advantage of an assembly incompetent mutant dimer, Cp149-Y132A, located in the inter-dimer interface. This mutation leads to changes in protein dynamics throughout the structure of the dimer as measured by hydrogen deuterium exchange mass spectrometry (HDX-MS). To further understand how the HBV capsid assembles, the homolog woodchuck HBV (WHV) capsid protein dimer (Cp) was used. WHV is more stable than HBV in HDX-MS and native mass spectrometry experiments. Because the WHV Cp assembles more rapidly into viral capsids than HBV, it was suspected that an increase in stability of the intra-dimer interface and/or in the contact region leads to increased assembly rates. The differences in dynamics when comparing HBV and human Cp149-Y132A as well as the differences in dynamics when comparing the HBV and WHV Cps allowed us to map an allosteric network within the HBV dimer. Through a careful comparison of structure, stability and dynamics using four different capsid protein dimers, we conclude that protein subunit dynamics regulate HBV capsid assembly.

Introduction

The human pathogen Hepatitis B Virus (HBV) chronically infects 257 million people worldwide and is responsible for around 800,000 deaths each year due to liver failure, hepatocellular carcinoma, and cirrhosis *(72)*. The mature virion consists of a protein studded outer membrane surrounding a nucleocapsid core which contains the gapped dsDNA genome and the reverse transcriptase *(73)*. The capsid assembles with T=4 (**Figure 3.1, panel A**) or T=3 symmetry which contain 240 or 180 copies of the core protein (Cp) respectively. T=4 particles are the most prevalent form and are the infectious particle *(74–79)*. The Cp consists of 183 amino acids and serves a variety of functions during the viral lifecycle *(80–84)*. The basic C-terminal domain is transiently exposed to the capsid exterior and is involved in intracellular trafficking through interactions with host cell proteins *(85–91)*. The first 149 amino acids of the Cp (Cp149) make up the assembly domain and are all that are needed for efficient capsid assembly *in vitro* (**Figure 3.1, panel B**) *(74,92–94)*.

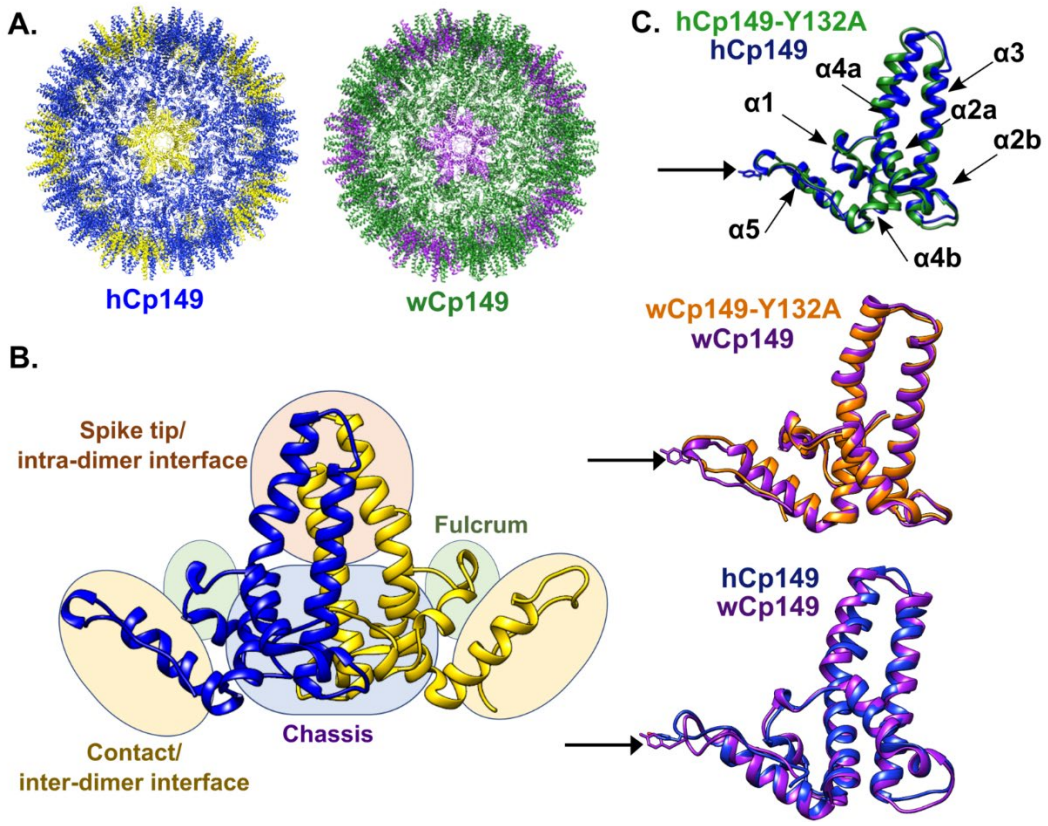

Figure 3.1. Structure of human and woodchuck hepatitis B viruses. A) T=4 capsid of human wild type HBV (left) T=4 capsid of the woodchuck wild type HBV (right). B) Subdomains of interest within the homodimer are shown. One monomer of the protein is displayed in blue while the other is yellow. C) Overlay of the mutant hCp149-Y132A crystal structure, which crystalized as a trimer of dimers, (green) over the wild type hCp149 crystal structure (blue) for human HBV in the capsid form (top, PDBIDs 3KXS and 1QGT, respectively). The black arrow indicates the location of the Y132A mutation. Overlay of the wCp149 in the capsid form (purple, PDBID 6EDJ) and wCp149-Y132A, which crystalized as a dimer of dimers, (orange, PDBID 6ECS) for the woodchuck HBV (middle). Overlay of the hCp149 (blue) with the wCp149 (purple, bottom). The sidechain of the mutated residue is displayed. Alpha helices are labeled.

Woodchucks chronically infected with woodchuck hepatitis B virus (WHV) develop hepatocellular carcinomas similar to humans infected with HBV *(95)*. Because of the similarities between WHV and HBV, WHV has been an important model for testing antiviral therapies *(95)*. The capsid proteins of these two viruses share 63% identity, with

the highest concentration of differences (23 of 55) in the intra-dimer interface (spike tip) (Supplemental Figure 1; Figure 1, panel C) *(96)*. Woodchuck Cp149 (wCp149) dimers assemble into about half T=4 capsids *in vitro*, with the other half forming particles larger than T=4 capsids *(97)*. While T=3 capsids are present in wCp149 assembly reactions, they appear less common than in hCp149 assembly reactions *(96)*.

Assembly rates of both hCp149 and wCp149 can be regulated by adjusting dimer concentration, pH, and ionic strength *(93,96,98,99)*. By maintaining a low dimer concentration, low ionic strength, and a near neutral pH, it is possible to study hCp149 and wCp149 proteins in as dimers *(98)*. The ability to study capsid building blocks free in solution, presents an opportunity to develop a deeper understanding of a regulated assembly mechanism. While the tertiary and quaternary structures of WHV and HBV capsids are very similar, the association energy differs. wCp149 assembles at lower ionic strength and temperature than hCp149 and assembly reactions carried out at the same ionic strength show that the formation of wCp149 capsids is more thermodynamically favorable *(96)*. WHV Cp149 assembles more rapidly than hCp149, with a substantially stronger association energy, and a high positive ΔC_p, which differs from hCp149's near zero ΔC_p *(96,100)*. The majority of primary sequence differences between WHV and HBV Cps occur in the two helices forming the intra-dimer interface. Nonetheless, these substitutions lead to a difference in assembly; suggesting that allostery may play a role in the assembly process.

An assembly incompetent mutant serves as a tool to connect aspects of protein structure and/or dynamics that modulate capsid assembly. Mutation of tyrosine at

position 132 to an alanine (Y132A) in either hCp149 or wCp149 renders the dimer assembly incompetent. The Y132A mutation is located in the inter-dimer contact region (**Figure 3.1, panel C**) *(101)*. In HBV, Y132 makes up 10% of the buried surface area when the dimer is assembled into capsids, however HBV Cp149-Y132A can co-assemble with WT hCp149 to form fragile capsids *(102)*. Therefore, it is capable of adopting an assembly competent conformation, though the mutation weakens dimer-dimer interactions. The structures of hCp149-Y132A and hCp149 are similar; however, it is somewhat surprising that the largest differences occur in the spike tip and intra-dimer interface, (**Figure 3.1, panel C**)*(101)* leading us to propose that changes in protein dynamics, both at the site of the mutation and areas distal to the mutation, may be the reason for the functional difference. If changes in dynamics control assembly of dimers into capsids, this suggests that the mutation alters the energy landscape. By comparing the differences in dynamics between hCp149-Y132A and WT hCp149, an understanding of whether these proteins sample different conformations at different rates can start to be determined.

It is suspected that there is an allosteric relationship between the spike tip, at the intra-dimer interface, and the contact region, at the inter-dimer interface, based on biophysical studies of hCp149, wCp149, hCp149-Y132A, and wCp149-Y132A *(100,101,103)*. We sought to provide new evidence for this relationship by directly measuring the intra-dimer interface stability and overall protein dynamics for each of these dimers. In this study, we used hydrogen deuterium exchange coupled to mass spectrometry (HDX-MS), native mass spectrometry (NMS), and differential scanning

fluorimetry (DSF) to investigate the biophysical differences between hCp149, wCp149, hCp149-Y132A, and wCp149-Y132A. wCp149 can more readily assemble into capsids than hCp149. This suggests that wCp149 possesses a conformer in its ensemble that is more readily poised for oligomerization. By comparing the hCp149 to the wCp149, we were able to determine that the spike tip and intra-dimer contact region of Cp149 are important functional regions. The assembly incompetent conformation of hCp149-Y132A shows a decrease in stability at the intra-dimer interface. The differences in stability and dynamics at the intra-dimer interface show that there is a connection between the contact region (where the mutation occurs) and this interface, which is a direct effect of the allosteric network that controls the hCp149 protein. This suggests that decreased intra-dimer stability destabilizes a key conformation in the energy landscape, making the free energy of assembly for hCp149-Y132A unfavorable.

Methods

Purification of HBV and WHV Dimers

Four expression plasmids, pET11a-hCp149, pET11a-hCp149-Y132A, pET11a-wCp149 and pET11a-wCp149-Y132A, were transformed into *E. coli* BL21(DE3) for protein expression. All proteins were purified based on a previously published protocol *(104)*. Specifically, each protein was purified as described: hCp149 *(93,104)*, hCp149-Y132A *(101)*, wCp149 *(96)*, and wCp149-Y132A *(100)*.

Differential Scanning Fluorimetry (DSF) of HBV and WHV Dimers

To determine the thermal stability of hCp149, hCp149-Y132A, wCp149, and wCp149-Y132A differential scanning fluorimetry was used. Reactions contained 3-8µg protein in 50 mM ammonium bicarbonate pH 7, as well as 2.5 µL of a 50x stock of SYPRO orange dye (Invitrogen) for a total reaction volume of 30 µL. The samples were loaded into a Rotor-Gene Q instrument (Qiagen) with SYPRO orange absorbance monitored at 570 nm as temperature was ramped from 25°C-95°C at 1°C/minute.

Native Mass Spectrometry (NMS) of HBV and WHV Dimers

Dimer stocks at 2 mg mL^{-1} were buffer exchanged into 10 mM ammonium acetate pH 7 using mini-dialysis cassettes (Thermo-Fisher) overnight at 4°C. Buffer exchanged dimers were then diluted 1:10 into 10 mM ammonium acetate pH 7. Buffer exchanged and diluted dimers were then loaded into in house prepared gold coated borosilicate capillaries prepared as outlined in Luo *et al.(105)* and analyzed using a Waters Synapt G2-*Si* with the following settings: capillary kV was set to 2.11, sampling cone: 148, source offset: 150, trap collision energy: 20.0-140V, transfer collision energy: 5.0V, source temperature: 60°C, and trap gas: 7.0 mL/minute. Trap collision energy was increased by 10V every 30 seconds from 20V to 140V. Data was analyzed using Waters MassLynx version 4.1Hydrogen-Deuterium Exchange Coupled to Mass Spectrometry (HDX-MS) of HBV and WHV Dimers

Hydrogen-Deuterium Exchange Coupled to Mass Spectrometry (HDX-MS) of HBV and WHV Dimers

Dimer (hCp149, hCp149-Y132A, wCp149, and wCp149-Y132A) stocks at 2 mg mL^{-1} were diluted 1:10 into deuterated reaction buffer (10 mM HEPES, pD 7.5 in D_2O). Control samples (time zero) were diluted into non-deuterated reaction buffer. At each time point (0,0.25, 3, 30, 180, and 1800 minutes) 10 µL of the reaction solution was removed and quenched by adding it to 60 µL of 0.75% formic acid (FA, Sigma) and 0.25 mg mL^{-1} porcine pepsin (Sigma) at pH 2.5 on ice. Each sample was digested for two minutes with vortexing every 30 seconds and then flash frozen in liquid nitrogen. Samples were stored in liquid nitrogen until liquid chromatography mass spectrometry analysis (LC-MS).

LC-MS analysis of HBV dimers was completed on a 1290 UPLC series chromatography stack (Agilent Technologies) coupled directly to a 6538 UHD Accurate-Mass QTOF LC/MS mass spectrometer (Agilent Technologies). Before electrospray−time-of-flight (ESI-TOF) analysis, peptides were separated on a reverse phase (RP) column (Phenomenex Onyx Monolithic C18 column, 100 × 2 mm) at 1 °C using a flow rate of 500 µL/min under the following conditions: 1.0 min, 5% B; 1.0−9.0 min, 5−45% B; 9.0−11.8 min, 45–95% B; 11.80−12.0 min, 5% B; solvent A = 0.1% FA (Sigma) in water (Thermo-Fisher) and solvent B = 0.1% FA in acetonitrile (ACN, Thermo-Fisher). Data were acquired at 2 Hz s^{-1} over the scan range 50−1700 m/z in positive mode. Electrospray settings were as follows: nebulizer set to 3.7 bar, drying gas at 8.0 L/min, drying temperature at 350 °C, and capillary voltage at 3.5 kV.

Data processing was carried out as outlined in *Berry et al.(106)* Briefly, peptides were identified using MassHunter Qualitative Analysis version 6.0 (Agilent Technologies), Peptide Analysis Worksheet (PAWs, ProteoMetrics LLC), and Peptide Shaker version 1.16.42 paired with Search GUI version 3.3.16 (Compomics)*(107,108)*. Peptide mapping to the structure was carried out as outlined in van Erp *et al.* and Berry *et al.(106,109)* Briefly, nested peptides were resolved. The amount of deuterium taken up per amide was determined by taking the amount of deuterium taken up by the peptide and dividing it by the number of amide hydrogens minus the two N-terminal amides. This amount of deuterium taken up per amide was then divided by the amount of deuterium taken up per amide for the 24-hour time point to determine the percentage of deuterium uptake. The percentage of deuterium uptake was compared between conditions and mapped to the HBV dimer structure.

Results

Differential Scanning Fluorimetry (DSF) of HBV and WHV Dimers

To determine if differences in the amino acid sequence at the spike tip or a mutation in the contact region alters the overall stability of the dimers, DSF was used. Thermal stability melting curves were generated for the dimers: hCp149, hCp149-Y132A, wCp149, and wCp149-Y132A. In DSF, the sample temperature is slowly raised while recording the fluorescence intensity of a dye that binds to hydrophobic pockets that become exposed as a protein unfolds *(110,111)*. hCp149 had a melting temperature of

68ºC (**Figure 3.2**), which is consistent with the melting temperature of 67°C for the reduced dimer observed by Wingfield *et al.* using differential scanning calorimetry*(94)*. The hCp149-Y132A dimer was not significantly different from hCp149 dimer. The melting temperatures of the two types of woodchuck dimer were lower than the human dimers by around 5°C. The similar melting temperatures between the wild type dimers and the mutant dimers indicates that the Y132A mutation has little effect on the overall thermal stability of the dimer. However, the changes in primary structure between WHV and HBV dimers do cause a shift in melting temperature. DSF measurements were repeated across a range of buffer conditions to assure that the differences in melting temperature are intrinsic properties of the dimers themselves and not an artifact of ionic strength or buffer conditions.

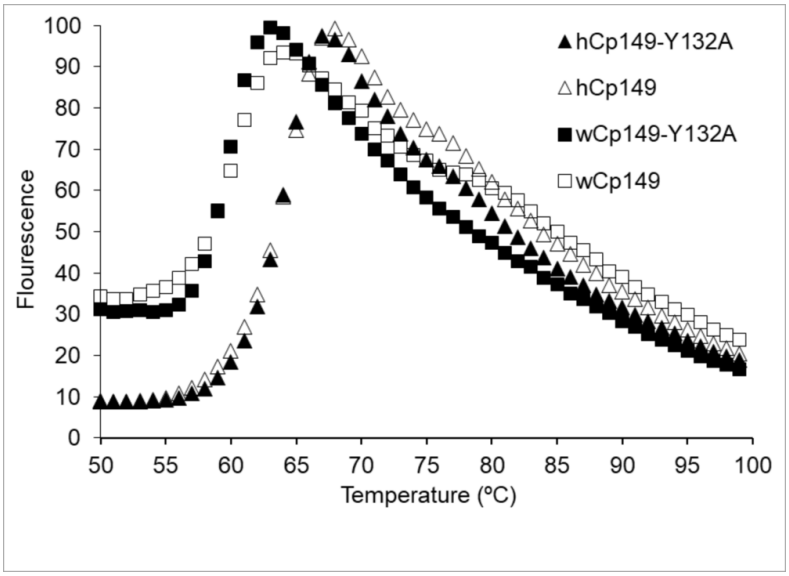

Figure 3.2. Differential scanning fluorimetry of HBV and WHV dimers. From this experiment it was determined that the two human Cp dimers have similar thermal denaturation temperatures which are higher than the two woodchuck Cp dimers.

Native Mass Spectrometry Probes Intra-Dimer Interface Stability

To further investigate the intra-dimer interface stability, a set of native mass spectrometry collisional induced unfolding experiments were performed on each of the dimers. Solutions containing dimers at 0.2 mg mL^{-1} in 10 mM ammonium acetate were analyzed by direct infusion, producing spectra that contained charge state distributions for both the monomer and the dimer (**Figure 3.3**). Low concentrations and low ionic strength conditions were used to eliminate the presence of multimers by limiting dimer-dimer interaction or non-specific clustering in the ionization process *(93,96,98,99,112)*. To probe the relative strength of the intra-dimer interface, the collision energy was increased in 10V increments from 20V to 140V while monitoring a specific dimer charge state (11+). As the energy was increased, the intensity of the peak for the 11+ ion, which only has contributions from the dimeric species, decreased indicating that the dimer was being broken apart into monomers in the mass spectrometer. When the collision energy exceeded 120V the intensity of the 11+ peak was significantly decreased for all dimers, allowing it to be used as a proxy for dimer stability.

The 11+ dimer peak intensity at 120V was compared to the initial dimer 11+ peak intensity at 20V (**Figure 3.3, panel C**). The hCp149-Y132A dimer was the least stable based on the changes in intensity from 20V to 120V followed by the hCp149, the wCp149-Y132A, and finally the wCp149. The Y132A mutation lowered the intra-dimer interface stability for both human and woodchuck dimers. The mutation is distal to the intra-dimer interface, yet, has a clear effect on interface stability. The woodchuck dimers had a stronger intra-dimer interface than the human dimers. This could be due to

differences in the amino acid sequence between the human and woodchuck dimers specifically in the intra-dimer interface.

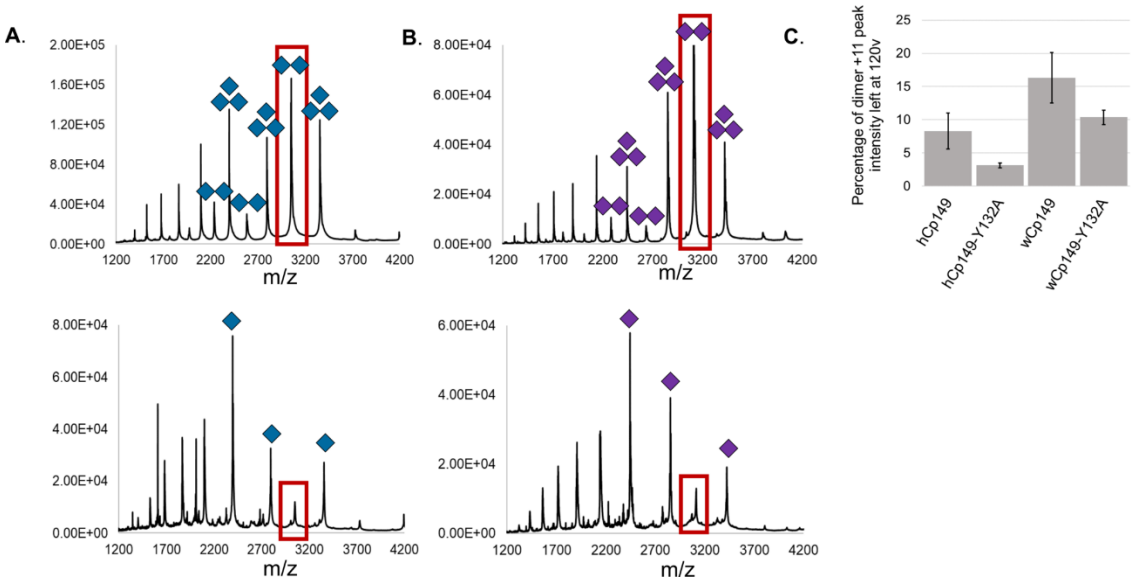

Figure 3.3 Native mass spectrometry of hCp149 and wCp149. Spectra at 20V collision energy (top) and 120V of collision energy (bottom) for human Cp149. A), woodchuck Cp149. B). The +11 dimer peak is indicated by a red box. As collision energy is increased, the intensity of the dimer decreased. Monomer peaks are indicated by one diamond, dimer peaks by two diamonds, and peaks that have contribution from both monomeric and dimeric species are indicated by three diamonds C) Percentage of the +11 dimer peak intensity left at 120V collision energy. Error bars are representative of the standard deviation of three experiments. Percentages of the dimer +11 peak left at other collision energies can be found in Supplemental Material Figure 5. The intensity of the +11 peak at 120V was divided by the intensity of the +11 peak at 20V. The intensity of the peak is proportional to how much of the dimer species is present in the sample. The +11 peak intensity is a measure of the relative stability of the dimers.

HDX-MS of wCp149 and hCp149 Dimers

Because the secondary and tertiary structure of WHV and HBV Cp149 dimers are highly similar *(100)*, we hypothesized that the differences in melting temperature, intra-dimer interface stability, and assembly rates arise from protein dynamics. To investigate the conformational flexibility between the two dimers, we used HDX-MS. Hydrogen

deuterium exchange measures the rate of exchange of the amide hydrogen on the protein backbone with solvent *(18,20)*. The exchange rate is inversely related to the stability of hydrogen bonding in the secondary and tertiary structure of the protein *(113)*.

For HDX-MS experiments, each dimer was diluted into deuterated buffer and allowed to exchange for specific lengths of time. Reaction conditions were optimized to maintain the dimeric state of each of the proteins tested. Exchange was quenched with acid followed by pepsin digestion and analysis by LC-MS. Pepsin digestion created a reproducible set of peptides that were used for analysis. The amount of deuterium taken up per peptide was determined at 0.25, 3, 30, 180, and 1800 minutes.

We found that the intra-dimer interface of the wCp149 spike tip takes up less deuterium than hCp149 dimer (**Figure 3.4**). That is, the hydrogen bonding network of the wCp149 spike tip has a more stable intra-dimer interface. To determine if amino acid sequence differences between human and woodchuck proteins in this region were responsible for the observed differences in exchange, we calculated the intrinsic rates of exchange from the sequences of the specific peptides. The observed rates of exchange at the 15 second time point for hCp149 (84-103) and wCp149 (84-102) differed by greater than three-fold while the intrinsic rates of exchange differed by less than 5%. This indicates that the observed differences are unlikely to be strictly a factor of sequence specific intrinsic exchange and rather are the result of changes in the stability of the hydrogen bonding network in the spike tip. In contrast, the fulcrum and inter-dimer contact regions (helix 5) had greater exchange in wCp149 than the hCp149 dimer. The calculated intrinsic rates of exchange in the contact region (helix 5) are similar, giving

confidence to the interpretation that this region is more dynamic in the wCp149 dimer than in the hCp149 dimer. However, the calculated intrinsic rate of exchange is different in the fulcrum region between the two proteins, so for the fulcrum it is not possible to confidently assign this to a change in protein stability and dynamics.

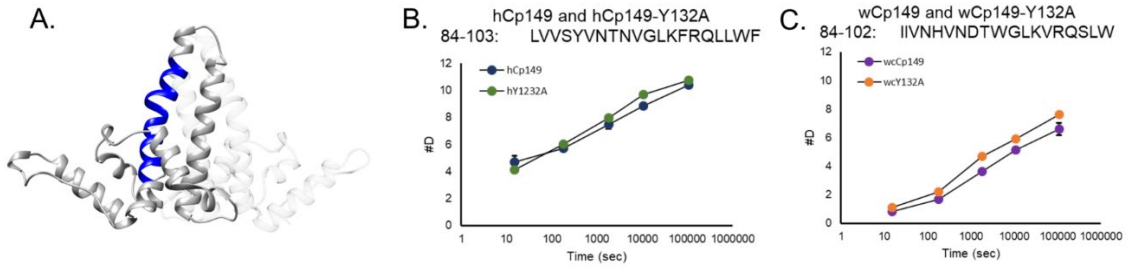

Figure 3.4. Deuterium uptake in the intra-dimer interface. between hCp149 and wCp149 at the intra-dimer interface. Helix 4a which is comprised of amino acid residues 79-109, is central to stability at the intra-dimer interface (A). The peptide 84-103 in hCp149 or 84-102 in wCp149 is shown in blue. At the three-hour time point, the human peptide takes up 8.8 deuterium (B) whereas the woodchuck peptide takes up 5.1 deuterium (C). Even though the hCp149 peptide is longer by one amino acid residue, the difference is statistically significant with a p-value < 0.01 at all time points. Calculations of the intrinsic rate of exchange are within 5% for the two peptides.

HDX-MS Analysis of Mutant Cp149

If an allosteric network connects the inter-dimer interface to the intra-dimer interface, we reasoned there should be differences in dynamics between the mutant Cp149-Y132A dimers and the wild type Cp149 dimers. In support of this hypothesis, we observed by HDX-MS that the Y132A mutation led to differences in exchange that propagated throughout the entire structure (**Figure 3.5, panel A**). Also, differences in deuterium uptake were not the same between the human and woodchuck dimers. In the HBV dimer, the site of the Y132A mutation was more flexible in the mutant and the chassis and fulcrum were more stable in the mutant. The spike tip showed more exchange

at earlier time points in the hCp149 dimer, but at three hours this trend switched and the hCp149-Y132A mutant took up more deuterium overall. This suggests that the spike has at least two populations of amides, fast and slow exchanging, that respond differently to the distal Y132A mutation.

In the wCp149 dimer, the residues near the Y132A mutation had an increase in exchange at early time points, but at later time points shows no difference in exchange when compared to wild type (**Figure 3.5, panel A**). This shows that the overall secondary structure of this region is not affected by the mutation at position 132, however, hydrogen bonding is less stable in the mutant implying that it is more dynamic than the wild type. The fulcrum and the chassis of the two forms of woodchuck have the same trends as the human forms of the dimer at early time points. At later time points, the chassis has more exchange in the wCp149-Y132A mutant indicating that this region has a less stable secondary structure in the mutant. The spike tip shows more exchange in the wCp149-Y132A mutant than in the wCp149 dimer indicating that the mutation destabilizes the hydrogen bond network on the other side of the protein. An important point is that there are changes in exchange throughout the HBV Cp dimer when the tyrosine at position 132 is mutated to an alanine, consistent with the idea that this is an allosterically regulated protein.

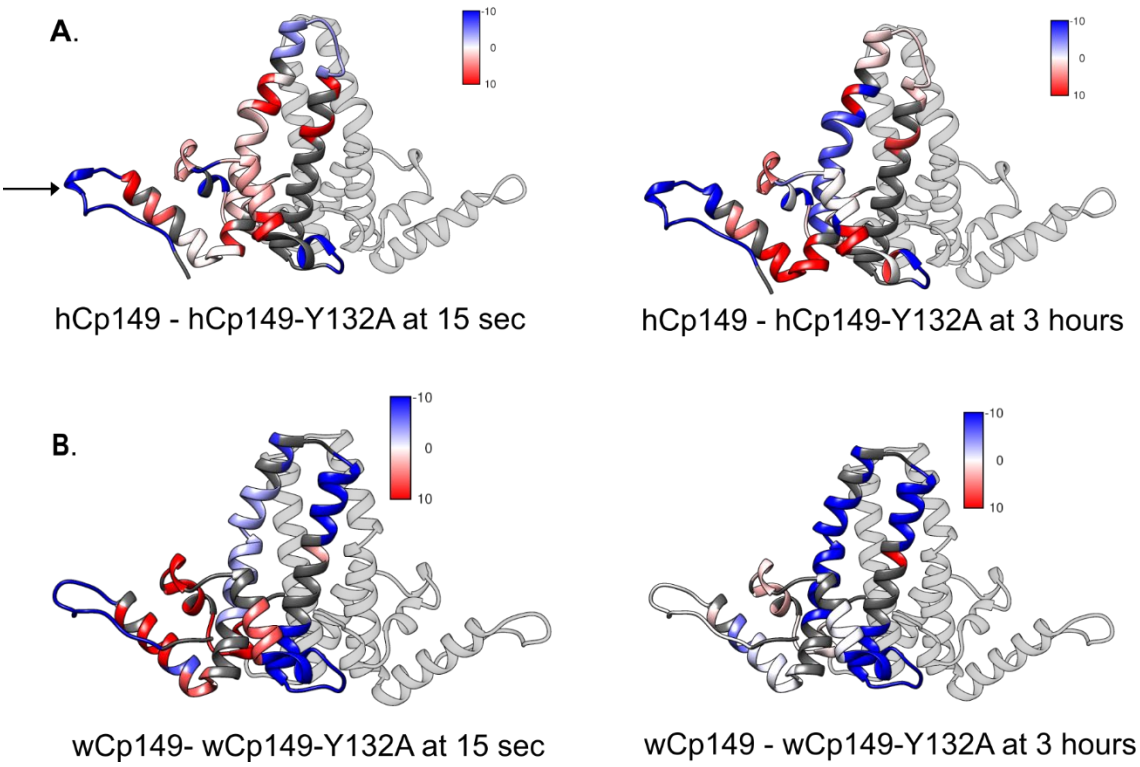

Figure 3.5. Mutation at residue 132 alters exchange across Cp149. A. The difference in the amount of deuterium taken up by the wild type and Cp149-Y132A human dimers at 15 seconds (left) and 3 hours (right). Areas where the wild type dimer takes up more deuterium than the Cp149-Y132A mutant dimer are displayed in red. Areas where the mutant Cp149-Y132A dimer takes up more deuterium than the wild type dimer are displayed in blue. Regions with no difference are displayed in white and regions that were not reported in this experiment are displayed in gray. B. The difference in the amount of deuterium taken up by the wild type and Cp149-Y132A woodchuck dimers at 15 seconds (left) and 3 hours (right). The location of the Y132A mutation is shown with an arrow in panel A.

Discussion

Biophysical, structural, and computational studies of the HBV capsid have helped to build a model in which protein dynamics and allostery are critical factors in the assembly and function of the T=4 particle. To date, a detailed model incorporating the behavior of assembly incompetent mutants is lacking. To understand the role of allostery

in the capsid assembly process of the HBV Cp149 protein, we compared the melting temperature, the strength of the intra-dimer interface, and protein dynamics of hCp149, hCp149-Y132A, wCp149, and wCp149-Y132A. These four dimers are nearly identical in structure, but perform differently in assembly assays. Both the hCp149-Y132A and wCp149-Y132A mutants are blocked from assembling into capsids. The Y132A mutation, located in the inter-dimer contact region, causes changes throughout the protein including stability of the intra-dimer interface. wCp149 assembles into capsids more rapidly than its human homolog. However, the proteins differ in sequence mainly in the spike tip region. This difference in sequence does not greatly affect the overall structure of the dimer, but it does cause changes to both the dynamics of the dimer and the stability of the intra-dimer interface.

Stability of the HBV Intra-Dimer Interface

The wCp149 and hCp149 dimers share 63% sequence identity and are structurally very similar *(96)* (**Figure 3.1**). Therefore, we were not expecting hCp149 to display 5ºC greater thermal transition than wCp149. Most sequence differences are localized to the spike tip. NMS and HDX-MS data show that the intra-dimer interface is more stable in wCp149 than in hCp149. This increase in stability may play a role in the ability of the wCp149 to assemble more readily into capsids than the hCp149. Our data also show that both the hCp149-Y132A and wCp149-Y132A mutations, which are well removed from the spike tip, none-the-less decreases stability at the intra-dimer interface. In both the human and the woodchuck dimers, the collision energy needed to disrupt the dimer in NMS was lower for the mutant (**Figure 3.3**). An increase in deuterium exchange was also

detected in the mutants in these regions at the three-hour time point (**Figure 3.5**). Of the four dimers, hCp149-Y132A was the least stable in NMS and had the most deuterium exchange in the spike region (**Figure 3.3C, Figure 3.4**), which is consistent with a less stable intra-dimer interface and more dynamic spike tip.

In Zhao *et al.,* it was determined that a decrease in stability at the intra-dimer interface caused by the mutation D78S reduced the rate of capsid assembly *(103)*. This mutation, located in the top of the intra-dimer interface, destabilizes helix 4a leading to a more flexible spike tip. This mutation slowed the assembly rate of the capsid showing that destabilization of the intra-dimer interface directly affects assembly kinetics. Together with our data, a model emerges that in order to have productive capsid assembly, the intra-dimer interface must be stable.

If increased stability in the spike tip causes the capsid to assemble more readily, the decrease in stability of this region in the hCp149-Y132A mutant should contribute to the inability of this mutant to assemble into capsids. However, there is a limit to the amount of stability that is beneficial to capsid assembly. Selzer *et al.* determined that if the intra-dimer interface is stabilized by a disulfide bond between the two C61 residues present in the interface, the rate of capsid assembly is slowed *(114)*. This trend in stabilization was also observed by Zhao *et al.* when investigating the oxidized wild-type human dimer and the oxidized D78S mutant dimer assembly at differing salt concentrations. These human dimers showed similar assembly kinetics to one another which were both slower than the wild-type hCp149 assembly kinetics, though the difference in rate of assembly decreased at higher ionic strengths *(103)*.Taken together,

these data indicate that there is a balance between stability and dynamics in this interface that must be maintained for productive capsid assembly *(103,114)*.

Inter-Dimer Contact Region of WHV and HBV

The contact region, which mediates dimer-dimer interactions, shows a complex pattern of deuterium exchange. The loop containing residue 132, located within the contact region, takes up more deuterium in both the hCp149-Y132A and wCp149-Y132A mutants than in the WT dimers at early time points (**Figure 3.5**). This implies that the mutation causes this region to be more flexible. In the α5 helix (**Figure 3.1C**), both the human and the woodchuck Cp149-Y132A mutants take up less deuterium than the respective WT dimers (**Figure 3.5**), as the helix becomes more structured upon mutation of Y132. Loss of structure in the loop of the contact region and a gain of structure in the α5 helix may contribute to the inability of the hCp149-Y132A and wCp149-Y132A mutants to assemble into capsids. The loop within the contact region shows similar deuterium exchange between the wCp149-Y132A and the hCp149-Y132A, however, there is more deuterium exchange in the wCp149-Y132A as compared to the hCp149-Y132A in the α5 helix. An increase in deuterium exchange is consistent with increased disorder in this region in the WHV Cp149-Y132A as compared to the HBV Cp149-Y132A. While we cannot be certain that the observed exchange rates in HDX were not partially caused by differences in intrinsic exchange rates due to changes in primary amino acid sequence, our results are supported by the observations made by Zhao *et al.* through limited proteolysis *(100)*. This is also consistent with Hilmer *et al.,* who measured the rate of proteolytic cleavage of the C-terminal region of the hCp149 protein

in free dimer and capsid. It was determined that the region near Y132 in the dimer becomes more folded in conditions that favor assembly *(115)*. This implies that stabilization of the contact region is necessary for productive assembly into capsids.

Comparison of Dynamics Between the hCp149 Free Dimer and hCp149 Capsid

In Bereszczak *et al.*, comparisons between the dynamics of the wild type HBV dimer free in solution and within the capsid show similar hydrogen deuterium exchange patterns in all regions except in the spike tip (residues 89-103) and the contact region (residues 122-140) which show less overall deuterium uptake in the capsid when compared to the free dimer *(116)*. In our data, the spike tip (intra-dimer interface) shows more deuterium incorporation in the hCp149-Y132A mutant than in the WT hCp149 dimer at later time points (**Figure 3.5, panel A**). The loss of stability (increased deuterium exchange) in this region may be detrimental to capsid assembly. In the inter-dimer contact region, hCp149-Y132A takes up more deuterium than hCp149 (**Figure 3.5A**), indicating that this could be another region where an increase in dynamics blocks capsid assembly. One interpretation of this is that the effect of the Y132A mutation on the spike tip and contact regions is to increase entropy. This must be compensated for during assembly, making the overall assembly process less thermodynamically favorable.

Our data show that changes in capsid assembly occur when the stability ad dynamics of the dimer change. The inter-dimer contact region has increased stability in the capsid and the capsid assembly-competent conformation. This seems logical based on the buried surface area between dimers in this region when assembled. The spike tip also shows an increase in stability in assembly competent dimers. The correlation of the

dynamics in the distal spike tip with assembly is characteristic of an allosteric network being present across the Cp149 dimer. Based on our dynamics data, this network spans the whole protein, connecting the inter-dimer interface (contact region) to the intra-dimer interface (spike tip).

An outcome of this work is that a deeper understanding of the capsid assembly process will provide targets for the further development of antivirals that disrupt the infection cycle of HBV. Our work highlights regions of the dimer where altered stability and dynamics lead to either enhanced or diminished capsid assembly. Current assembly directed HBV antivirals act by increasing buried surface area in the inter-dimer contact region generating large aberrant structures *(117)*. An important outcome from our work is that the spike tip or intra-dimer interface represents another target for assembly disruptors.

Conclusions

Dynamic proteins exist as an ensemble of conformers in solution. Modulation of the energy landscape of the ensemble, and thus the relative population, through changes in conformation or dynamics, is the business of allostery *(12)*. To understand how HBV Cp149 assembles into capsids, we have worked to bring allosteric features into the light. The assembly competent conformer of HBV has a stable, but not too stable, intra-dimer interface and a stable inter-dimer interface. Our data strongly suggest that the Y132A mutation alters the ensemble of conformations present in solution. This implies that the hCp149-Y132A dimer does not sample the assembly active conformation. However, the WHV Cp149, which is structurally similar to the HBV Cp149, assembles more rapidly

implying that it more frequently samples a capsid assembly competent conformer. HBV dimer is an excellent example of why structural models alone can be insufficient when trying to understand and predict functional outcomes of mutations. By using information about the energy landscape and dynamics of a protein or complex, even sophisticated mechanisms can be elucidated.

References

(72) CDA Foundation. CDA Foundation's Polaris Observatory https://cdafound.org/polaris/ (accessed Apr 9, 2020).

(73) Seeger, C.; Mason, W. S. Molecular Biology of Hepatitis B Virus Infection. *Virology* **2015**, *479–480*, 672–686. https://doi.org/10.1016/j.virol.2015.02.031.

(74) Zlotnick, A.; Cheng, N.; Conway, J. F.; Booy, F. P.; Steven, A. C.; Stahl, S. J.; Wingfield, P. T. Dimorphism of Hepatitis B Virus Capsids Is Strongly Influenced by the C-Terminus of the Capsid Protein. *Biochemistry* **1996**, *35* (23), 7412–7421. https://doi.org/10.1021/bi9604800.

(75) Bringas, R. Folding and Assembly of Hepatitis B Virus Core Protein: A New Model Proposal. *J. Struct. Biol.* **1997**, *118* (3), 189–196. https://doi.org/10.1006/jsbi.1997.3846.

(76) Crowther, F. A.; Kiselev, N. A.; Bottcher, B.; Berriman, J. A.; Borisova, G. P.; Ose, S. V; Pumpens, P. Three-Dimensional Structure of Hepatitis B Virus Core Particles Determined by Electron Cryomicroscopy. *Cell* **1994**, *77*, 943–950. https://doi.org/10.1016/0092-8674(94)90142-2.

(77) Singh, S.; Zlotnick, A. Observed Hysteresis of Virus Capsid Disassembly Is Implicit in Kinetic Models of Assembly. *J. Biol. Chem.* **2003**, *278* (20), 18249–18255. https://doi.org/10.1074/jbc.M211408200.

(78) Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W. The Crystal Structure of the Human Hepatitis B Virus Capsid. **1999**, *3*, 771–780.

(79) Stannard, L. M.; Hodgkiss, M. Morphological Irregularities in Dane Particle Cores. *J. Gen. Virol.* **1979**, *45* (2), 509–514. https://doi.org/10.1099/0022-1317-45-2-509.

(80) Gazina, E. V.; Fielding, J. E.; Lin, B.; Anderson, D. A. Core Protein

Phosphorylation Modulates Pregenomic RNA Encapsidation to Different Extents in Human and Duck Hepatitis B Viruses. *J. Virol.* **2000**, *74* (10), 4721–4728. https://doi.org/10.1128/jvi.74.10.4721-4728.2000.

(81) Porterfield, J. Z.; Dhason, M. S.; Loeb, D. D.; Nassal, M.; Stray, S. J.; Zlotnick, A. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. *J. Virol.* **2010**, *84* (14), 7174–7184. https://doi.org/10.1128/jvi.00586-10.

(82) Freund, S. M. V.; Johnson, C. M.; Jaulent, A. M.; Ferguson, N. Moving towards High-Resolution Descriptions of the Molecular Interactions and Structural Rearrangements of the Human Hepatitis B Core Protein. *J. Mol. Biol.* **2008**, *384* (5), 1301–1313. https://doi.org/10.1016/j.jmb.2008.10.020.

(83) Alexander, C. G.; Jurgens, M. C.; Shepherd, D. A.; Freund, S. M. V.; Ashcroft, A. E.; Ferguson, N. Thermodynamic Origins of Protein Folding, Allostery, and Capsid Formation in the Human Hepatitis B Virus Core Protein. *Proc. Natl. Acad. Sci.* **2013**, *110* (30), E2782–E2791. https://doi.org/10.1073/pnas.1308846110.

(84) Seitz, S.; Urban, S.; Antoni, C.; Böttcher, B. Cryo-Electron Microscopy of Hepatitis B Virions Reveals Variability in Envelope Capsid Interactions. *EMBO J.* **2007**, *26* (18), 4160–4167. https://doi.org/10.1038/sj.emboj.7601841.

(85) Li, H. C.; Huang, E. Y.; Su, P. Y.; Wu, S. Y.; Yang, C. C.; Lin, Y. S.; Chang, W. C.; Shih, C. Nuclear Export and Import of Human Hepatitis B Virus Capsid Protein and Particles. *PLoS Pathog.* **2010**, *6* (10). https://doi.org/10.1371/journal.ppat.1001162.

(86) Kann, M.; Schmitz, A.; Rabe, B. Intracellular Transport of Hepatitis B Virus. *World J. Gastroenterol.* **2007**, *13* (1), 39–47. https://doi.org/10.3748/wjg.v13.i1.39.

(87) Chen, C.; Wang, J. C. Y.; Pierson, E. E.; Keifer, D. Z.; Delaleau, M.; Gallucci, L.; Cazenave, C.; Kann, M.; Jarrold, M. F.; Zlotnick, A. Importin β Can Bind Hepatitis B Virus Core Protein and Empty Core-Like Particles and Induce Structural Changes. *PLoS Pathog.* **2016**, *12* (8), 1–25. https://doi.org/10.1371/journal.ppat.1005802.

(88) Selzer, L.; Kant, R.; Wang, J. C. Y.; Bothner, B.; Zlotnick, A. Hepatitis B Virus Core Protein Phosphorylation Sites Affect Capsid Stability and Transient Exposure of the C-Terminal Domain. *J. Biol. Chem.* **2015**, *290* (47), 28584–28593. https://doi.org/10.1074/jbc.M115.678441.

(89) Yeh, C. T.; Liaw, Y. F.; Ou, J. H. The Arginine-Rich Domain of Hepatitis B Virus Precore and Core Proteins Contains a Signal for Nuclear Transport. *J. Virol.* **1990**, *64* (12), 6141–6147.

(90) Liao, W.; Ou, J. H. Phosphorylation and Nuclear Localization of the Hepatitis B Virus Core Protein: Significance of Serine in the Three Repeated SPRRR Motifs. *J. Virol.* **1995**, *69* (2), 1025–1029.

(91) Kann, M.; Sodeik, B.; Vlachou, A.; Gerlich, W. H.; Helenius, A. Phosphorylation-Dependent Binding of Hepatitis B Virus Core Particles to the Nuclear Pore Complex. *J. Cell Biol.* **1999**, *145* (1), 45–55. https://doi.org/10.1083/jcb.145.1.45.

(92) Gallina, A.; Bonelli, F.; Zentilin, L.; Rindi, G.; Muttini, M.; Milanesi, G. A Recombinant Hepatitis B Core Antigen Polypeptide with the Protamine-like Domain Deleted Self-Assembles into Capsid Particles but Fails to Bind Nucleic Acids. *J. Virol.* **1989**, *63* (11), 4645–4652.

(93) Ceres, P.; Zlotnick, A. Weak Protein-Protein Interactions Are Sufficient to Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry* **2002**, *41* (39), 11525–11531. https://doi.org/10.1021/bi0261645.

(94) Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C. Hepatitis Core Antigen Produced in Escherichia Coli: Subunit Composition, Conformation Analysis, and in Vitro Capsid Assembly. *Biochemistry* **1995**, *34* (15), 4919–4932. https://doi.org/10.1021/bi00015a003.

(95) Tennant, B. C.; Gerin, J. L. The Woodchuck Model of Hepatitis B Virus Infection. *ILAR J.* **2001**, *42* (2), 89–102. https://doi.org/10.1093/ilar.42.2.89.

(96) Kukreja, A. A.; Wang, J. C.-Y.; Pierson, E.; Keifer, D. Z.; Selzer, L.; Tan, Z.; Dragnea, B.; Jarrold, M. F.; Zlotnick, A. Structurally Similar Woodchuck and Human Hepadnavirus Core Proteins Have Distinctly Different Temperature Dependences of Assembly. *J. Virol.* **2014**, *88* (24), 14105–14115. https://doi.org/10.1128/jvi.01840-14.

(97) Pierson, E. E.; Keifer, D. Z.; Kukreja, A. A.; Wang, J. C. Y.; Zlotnick, A.; Jarrold, M. F. Charge Detection Mass Spectrometry Identifies Preferred Non-Icosahedral Polymorphs in the Self-Assembly of Woodchuck Hepatitis Virus Capsids. *J. Mol. Biol.* **2016**, *428* (2), 292–300. https://doi.org/10.1016/j.jmb.2015.06.019.

(98) Asor, R.; Selzer, L.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Assembly Reactions of Hepatitis B Capsid Protein into Capsid Nanoparticles Follow a Narrow Path through a Complex Reaction Landscape. *ACS Nano* **2019**. https://doi.org/10.1021/acsnano.9b00648.

(99) Asor, R.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Rapidly Forming Early Intermediate Structures Dictate the Pathway of Capsid Assembly. *J. Am. Chem. Soc.* **2020**, *142* (17), 7868–7882. https://doi.org/10.1021/jacs.0c01092.

(100) Zhao, Z.; Wang, J. C.-Y.; Gonzalez-Gutierrez, G.; Venkatakrishnan, B.; Asor, R.;

Khaykelson, D.; Raviv, U.; Zlotnick, A. Structural Differences between the Woodchuck Hepatitis Virus Core Protein in the Dimer and Capsid States Are Consistent with Entropic and Conformational Regulation of Assembly. *J. Virol.* **2019**, *93* (14), 1–16. https://doi.org/10.1128/jvi.00141-19.

(101) Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with a Role for Allostery in Virus Assembly. *J. Virol.* **2010**, *84* (3), 1607–1615. https://doi.org/10.1128/JVI.02033-09.

(102) Bourne, C. R.; Katen, S. P.; Fulz, M. R.; Packianathan, C.; Zlotnick, A. A Mutant Hepatitis b Virus Core Protein Mimics Inhibitors of Icosahedral Capsid Self-Assembly. *Biochemistry* **2009**, *48* (8), 1736–1742. https://doi.org/10.1021/bi801814y.

(103) Zhao, Z.; Wang, J. C.-Y.; Segura, C. P.; Hadden-Perilla, J. A.; Zlotnick, A. The Integrity of the Intradimer Interface of the Hepatitis B Virus Capsid Protein Dimer Regulates Capsid Self-Assembly. *Submitted* **2020**.

(104) Zlotnick, A.; Lee, A.; Bourne, C. R.; Johnson, J. M.; Paul, L. In Vitro Screening for Molecules That Affect Virus Capsid Assembly. *Nat. Protoc.* **2007**, *2* (3), 490–498. https://doi.org/10.1016/j.bbi.2008.05.010.

(105) Luo, M. L.; Jackson, R. N.; Denny, S. R.; Tokmina-Lukaszewska, M.; Maksimchuk, K. R.; Lin, W.; Bothner, B.; Wiedenheft, B.; Beisel, C. L. The CRISPR RNA-Guided Surveillance Complex in Escherichia Coli Accommodates Extended RNA Spacers. *Nucleic Acids Res.* **2016**, *44* (15), 7385–7394. https://doi.org/10.1093/nar/gkw421.

(106) Berry, L.; Patterson, A.; Pence, N.; Peters, J.; Bothner, B. Hydrogen Deuterium Exchange Mass Spectrometry of Oxygen Sensitive Proteins. *Bio-Protocol* **2018**, *8* (6), 1–16. https://doi.org/10.21769/BioProtoc.2769.

(107) Vaudel, M.; Burkhart, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker Enables Reanalysis of MS-Derived Proteomics Data Sets: To the Editor. *Nat. Biotechnol.* **2015**, *33* (1), 22–24. https://doi.org/10.1038/nbt.3109.

(108) Barsnes, H.; Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J. Proteome Res.* **2018**, *17* (7), 2552–2555. https://doi.org/10.1021/acs.jproteome.8b00175.

(109) Van Erp, P. B. G.; Patterson, A.; Kant, R.; Berry, L.; Golden, S. M.; Forsman, B. L.; Carter, J.; Jackson, R. N.; Bothner, B.; Wiedenheft, B. Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. *ACS Chem. Biol.* **2018**, *13* (2), 481–490.

https://doi.org/10.1021/acschembio.7b00649.

(110) Rayaprolu, V.; Kruse, S.; Kant, R.; Movahed, N.; Brooke, D.; Bothner, B. Fluorometric Estimation of Viral Thermal Stability. *Bio-Protocol* **2014**, *4* (15). https://doi.org/10.21769/bioprotoc.1199.

(111) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–2221. https://doi.org/10.1038/nprot.2007.321.

(112) Chen, F.; Gulbakan, B.; Weidmann, S.; Fagerer, S. R.; Ibanez, A. J.; Zenobi, R. Applying Mass Spectrometry to Study Non-Covalent Biomolecule Complexes. *Mass Spectrom. Rev.* **2015**, *35* (1), 48–70. https://doi.org/10.1002/mas.21462.

(113) Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. Protein Dynamics Viewed by Hydrogen Exchange. *Protein Sci.* **2012**, *21* (7), 996–1005. https://doi.org/10.1002/pro.2081.

(114) Selzer, L.; Katen, S. P.; Zlotnick, A. The Hepatitis B Virus Core Protein Intradimer Interface Modulates Capsid Assembly and Stability. *Biochemistry* **2014**, *53* (34), 5496–5504. https://doi.org/10.1021/bi500732b.

(115) Hilmer, J. K.; Zlotnick, A.; Bothner, B. Conformational Equilibria and Rates of Localized Motion within Hepatitis B Virus Capsids. *J. Mol. Biol.* **2008**, *375* (2), 581–594. https://doi.org/10.1016/j.jmb.2007.10.044.

(116) Bereszczak, J. Z.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Assessment of Differences in the Conformational Flexibility of Hepatitis B Virus Core-Antigen and e-Antigen by Hydrogen Deuterium Exchange-Mass Spectrometry. *Protein Sci.* **2014**, *23* (7), 884–896. https://doi.org/10.1002/pro.2470.

(117) Stray, S. J.; Bourne, C. R.; Punna, S.; Lewis, W. G.; Finn, M. G.; Zlotnick, A. A Heteroaryldihydropyrimidine Activates and Can Misdirect Hepatitis B Virus Capsid Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (23), 8138–8143. https://doi.org/10.1073/pnas.0409732102.

CHAPTER FOUR

CONFORMATIONAL DYNAMICS OF DNA BINDING AND CAS3 RECRUITMENT BY THE CRISPR RNA-GUIDE CASCADE COMPLEX

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Paul van Erp

Contributions: Designed the research, performed experiments, analyzed the data, and wrote the paper.

Author: Angela Patterson

Contributions: Performed experiments, analyzed the data, and wrote the paper.

Co-Author: Ravi Kant

Contributions: Designed the research and analyzed the data.

Co-Author: Luke Berry

Contributions: Analyzed the data and wrote the paper.

Co-Author: Sarah M. Golden

Contributions: Performed experiments.

Co-Author: Brittney L. Forsman

Contributions: Performed experiments.

Co-Author: Joshua Carter

Contributions: Analyzed the data.

Co-Author: Ryan N. Jackson

Contributions: Analyzed the data.

Co-Author: Brian Bothner

Contributions: Designed the research, analyzed the data, and wrote the paper.

Co-Author: Blake Wiedenheft

Contributions: Designed the research, analyzed the data, and wrote the paper.

Manuscript Information

Paul B.G. van Erp, Angela Patterson, Ravi Kant, Luke Berry, Sarah M. Golden, Brittney L. Forsman, Joshua Carter, Ryan N. Jackson, Brian Bothner, and Blake Wiedenheft

ACS Chemical Biology

Status of Manuscript:
____ Prepared for submission to a peer-reviewed journal
____ Officially submitted to a peer-reviewed journal
____ Accepted by a peer-reviewed journal
__x_ Published in a peer-reviewed journal

ACS Chemical Biology
Volume 13, Issue 2
DOI: 10.1021/acschembio.7b00649

Abstract

Bacteria and archaea rely on CRISPR (clustered regularly interspaced short palindromic repeats) RNA-guided adaptive immune systems for sequence specific elimination of foreign nucleic acids. In Escherichia coli, short CRISPR-derived RNAs (crRNAs) assemble with Cas (CRISPR-associated) proteins into a 405-kilodalton multisubunit surveillance complex called Cascade (CRISPR-associated complex for antiviral defense). Cascade binds foreign DNA complementary to the crRNA guide and recruits Cas3, a trans-acting nuclease-helicase required for target degradation. Structural models of Cascade have captured static snapshots of the complex in distinct conformational states, but conformational dynamics of the 11-subunit surveillance complex have not been measured. Here, we use hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) to map conformational dynamics of Cascade onto the three-dimensional structure. New insights from structural dynamics are used to make functional predictions about the mechanisms of the R-loop coordination and Cas3 recruitment. We test these predictions in vivo and in vitro. Collectively, we show how mapping conformational dynamics onto static 3D-structures adds an additional dimension to the functional understanding of this biological machine.

Introduction

The adaptive immune system in *Escherichia coli* (type I-E) consists of eight *cas* genes flanked by a CRISPR locus (**Figure 4.1A**). Five of the *cas* genes encode proteins ($Cse1_1$, $Cse2_2$, $Cas7_6$, $Cas5_1$, $Cas6e_1$) that assemble with a 61-nucleotide

CRISPR-derived RNA (crRNA) into a seahorse-shaped complex called Cascade (CRISPR-associated complex for antiviral defense), with subunits that resemble a head, backbone, belly, and tail (**Figure 4.1B**) *(118–121)*. Cascade is a crRNA-guided surveillance complex that identifies foreign DNA and recruits a trans-acting nuclease-helicase (i.e. Cas3) for target degradation *(122–124)*.

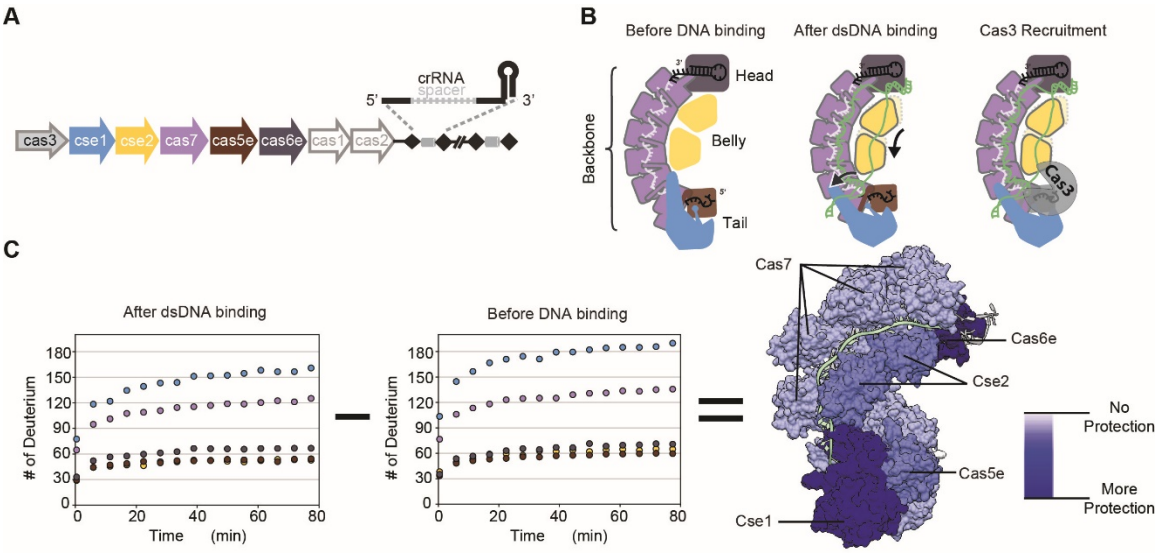

Figure 4.1 Binding of dsDNA increases the stability of Cascade. (**A**) The type 1-E CRISPR system of *E. coli* consists of eight *cas* genes and a CRISPR locus. Five *cas* genes (colored) code for proteins that assemble with a crRNA into the Cascade complex. (**B**) Schematic representation of three stages of Cascade during the interference process. Subunits colored according to Figure 1A. (**C**) Deuterium uptake curves for the individual subunits of Cascade. Error bars show the standard deviation (n=3) and are smaller than the circular symbols used to plot the data. Both the unbound and dsDNA bound Cascade complexes were incubated in deuterated buffer for 0.5 to 77.5 minutes. Colors of the data points on the uptake curves correspond with the subunits of Cascade as colored in Figure 1A. Differences in deuterium uptake were determined by subtracting deuterium uptake of the unbound complex from the dsDNA bound complex at 77.5 minutes. The percentage change in uptake was mapped onto the dsDNA bound structure of Cascade (5H9F). Increasing blue hue indicates greater protection (i.e., less deuterium uptake) in the dsDNA bound form. Overall, the dsDNA bound form of Cascade takes up less deuterium and is more protected.

Target DNA recognition by Cascade relies on protein-mediated recognition of a three-nucleotide PAM (protostpacer adjacent motif) and a protospacer sequence that is complementary to the crRNA-guide *(118,125–127)*. Structures of *E. coli* Cascade before binding DNA *(120,121,128)*, bound to ssDNA *(129)*, bound to a partially duplexed dsDNA target *(125)*, and recent structures of Cascade from *Thermobifida fusca* bound to a dsDNA target *(130)*, provide snapshots of the complex poised at different stages of antiviral defense. Collectively, these structures explain how DNA binding triggers a conformational rearrangement in which the two belly subunits (i.e. Cse2) slide along the ridged Cas7 backbone via a molecular relay that relies on a series of conserved arginine residues *(128)*. One of the Cse2 subunits abuts the C-terminal domain (CTD) of Cse1 (also called Cas8e), and the DNA-induced transition of Cse2 is thought to shove the CTD of Cse1 between the two strands of the DNA duplex *(120,121,125,129)*. While the CTD of Cse1 appears to function like a molecular pry bar that helps separate the two strands of the DNA duplex, three structural features (i.e. lysine finger, glycine rich loop and glutamine wedge) in the N-terminal domain of Cse1 are involved in PAM recognition *(125)*. PAM recognition relies on sequence- and shaped-based recognition of the minor groove, resulting in a diverse combination of different PAMs that are recognized with varying affinities *(124,125,131,132)*.

While a variety of PAMs support target binding, different PAM and protospacer combinations have been shown to elicit distinct biological outcomes *(131,133,134)*. Binding to a limited number of PAMs elicits an “interference response”, which is characterized by recruitment of a nuclease active form of Cas3 that processively degrades

the target *(124,134–137)*. Alternatively, some PAMs or mutations at specific positions in the protospacer, reduce interference to levels below what is required for clearance of plasmids or virus elimination, and trigger a "priming response." Priming is a positive feedback mechanism characterized by the rapid integration of new fragments of foreign DNA into the CRISPR locus, which are used to restore immunity *(131,133,134,138)*. Priming relies on all the of the CRISPR machinery (Cascade, Cas1, Cas2, and Cas3) in addition to a partial complementary target.

Recently, Förster resonance energy transfer (FRET) studies have been used to monitor the conformational state of Cascade in association with targets that elicit priming or interference. Xue *et al.* used structures of Cascade to position FRET dyes at specific locations in the tail of Cascade to monitor the conformational state of Cse1 in response to binding DNA targets that elicit priming or interference *(136,139)*. These studies demonstrated that Cascade adopts distinct conformational states based on sequence characteristics of the dsDNA target and that the conformational state of the Cse1 tail subunit correlates with the relative rates of interference and priming. While structure guided placement of FRET dyes provides a powerful means for measuring distances between two selected landmarks, this approach cannot provide a global understanding of the complex conformational dynamics of a large multi-subunit machine like Cascade.

Rather than examining movements of specific locations on Cascade, here we use hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) to investigate protein interactions and dynamics of Cascade before and after DNA binding. HDX measures the stability of hydrogen bond networks and solvent accessibility on global and

local scales *(140,141)*, enabling a mechanistic analysis of coordinated changes associated with target binding. To our knowledge, this is the first application of HDX-MS to a ribonucleoprotein of this complexity (i.e., 5 unique proteins totaling >1,450 amino acid residues). Peptide-resolution HDX-MS experiments reveal residues predicted to be involved in binding the displaced DNA strand (i.e. R-loop) and regions of the Cse1 subunit that are part of the Cas3 binding site. We test the function of these residues using *in vivo* and *in vitro* assays and show that HDX-guided mutations at specific locations eliminate Cas3-mediated DNA degradation without altering Cascade's ability to bind dsDNA.

Results

Structural studies have captured snapshots of Cascade before and after target binding, but these static structures provide limited information about the conformational dynamics of the complex, which are important for function *(120,121,125,129)*. To measure global changes in the stability and dynamics of Cascade upon target binding, we used intact protein HDX-MS to measure HD-exchange in all 11-protein subunits of Cascade, both before and after binding DNA. We diluted unbound Cascade and Cascade bound to a 72-base pair dsDNA target containing a 3′-TTC-5′ PAM into deuterated buffer and then analyzed both complexes by Liquid Chromatography Mass Spectrometry (LCMS) over 15 time points (**Figure 4.1**). At each time point, dsDNA bound Cascade exhibited less deuterium uptake than the unbound form, though the difference was less pronounced at later time points. Overall, the intact protein HDX-MS revealed a net stabilization of each protein subunit in the complex upon dsDNA binding (**Figure 4.1C**).

This stabilization was further confirmed using differential scanning fluorimetry, which showed that Cascade bound to a dsDNA target has a melting temperature that is approximately 10°C above the melting temperature of unbound Cascade.

Having established that DNA binding globally decreases deuterium uptake and increases stability, we set out to map these changes to specific regions of the complex. To accomplish this, we digested Cascade and dsDNA-bound Cascade with pepsin (i.e., a low specificity protease), prior to LCMS. Deuterium uptake curves were generated by quenching the HDX reaction after 0, 0.16, 0.5, 2.5, 5.0, 10, 30 and 180 minutes. 255 peptides, representing more than 90% of all Cascade protein sequences, were analyzed. Parent spectra (MS1) were used to determine the number of deuterons incorporated per peptide, and the difference in deuteron incorporation between peptides isolated from dsDNA bound and unbound states of Cascade were used to calculate changes in HD-exchange. Overall, 185 of the 255 peptides showed increased deuterium incorporation over the time course in either or both complexes, while 70 peptides did not. After binding dsDNA, 22 of the 255 peptides increased deuterium incorporation, while 119 peptides had decreased deuterium incorporation compared to the unbound complex.

HD-exchange profiles are not uniformly distributed across the complex or within an individual subunit. By analyzing the distribution of deuterium uptake over the time course, changes in the stability and dynamics of hydrogen bond networks upon dsDNA binding can be determined. Early time points report on changes in solvent accessibility upon dsDNA binding, whereas later time points provide information on the stability of hydrogen bonding networks and the participation of peptide backbone amides in protein

secondary structures that are protected from deuterium exchange because of noncovalent interactions *(142)*. Regions of the complex that displayed significant protection upon dsDNA binding (i.e. decreased deuterium exchange) were investigated as potential DNA binding sites (**Figure 4.2A,** blue). We also analyzed peptides that assimilate more deuterium (i.e. become less protected) after binding DNA (**Figure 4.2B,** red). An increase in exchange indicates greater solvent exposure or dynamics. These regions are anticipated to have functional roles, such as docking sites for the nuclease-helicase Cas3.

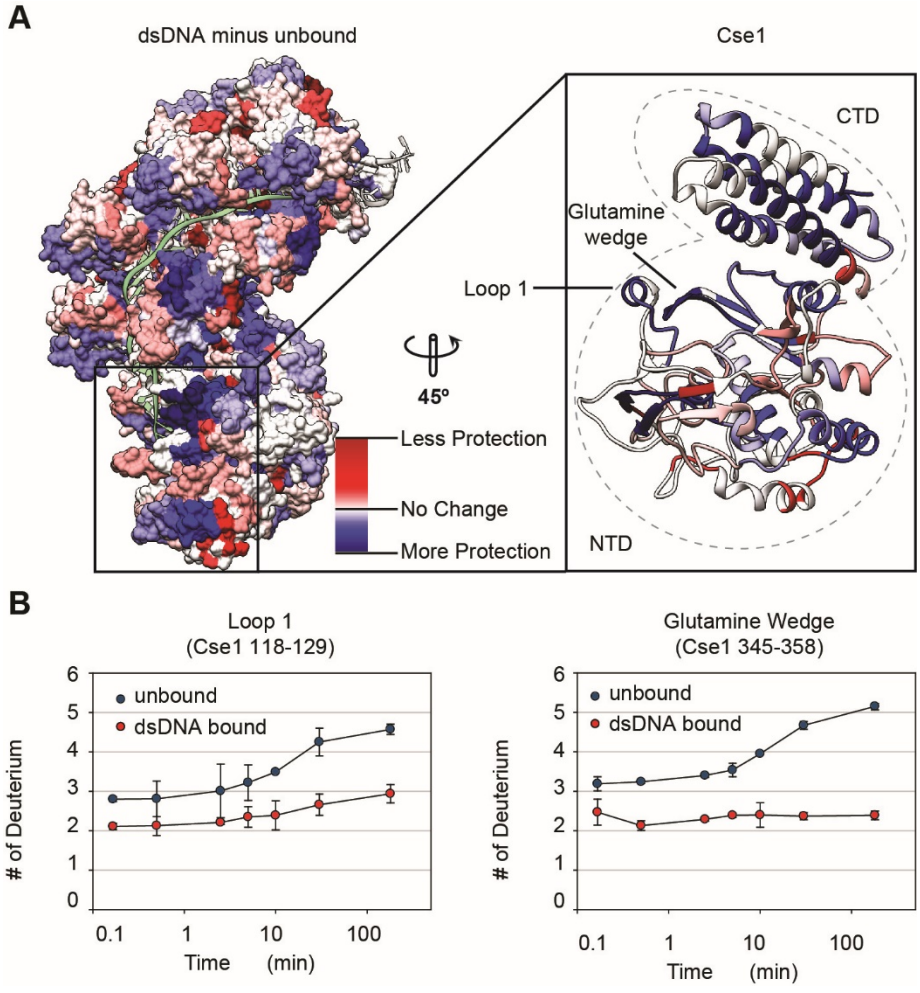

Figure 4.2. Peptide-resolution mapping of deuterium uptake reveals areas of Cascade that become more or less protected upon dsDNA binding. (A) The percentage difference in deuterium uptake between the unbound and dsDNA bound complex for individual peptides at the 180-minute time point was mapped onto the dsDNA bound structure (5H9F). Blue indicates less deuterium incorporation in the dsDNA bound form (protection) while the red indicates more deuterium incorporation in the dsDNA bound form (de-protection). White indicates regions where there is no difference in deuterium uptake or peptides where no data is available. The inset shows the ribbon structure of the Cse1 subunit with the glutamine wedge and Loop 1 showing increased protection. The Cse1 subunit consists of a four-helix bundle (CTD) and globular domain (NTD). (B) Deuterium uptake curves show increased protection in the dsDNA bound form for peptides in Loop 1 (p-value<0.01 for all time points) and the glutamine wedge (p-value<0.01 for all time points after 2.5 minutes). Error bars represent the standard deviation of three replicates.

DNA Binding

To validate the HDX data we examined regions of the complex known to interact with DNA. The expectation is that DNA binding will stabilize the peptide backbone and physically occlude these regions from HD-exchange. Recognition of the PAM sequence takes place in the tail of the complex and involves three structural elements in the Cse1 subunit: lysine finger, glycine loop, and glutamine wedge *(125)*. The glutamine wedge (Cse1 amino acids 352–358) forms a beta-hairpin that is positioned (i.e. wedged) between the complementary and non-complementary strands of the DNA target (**Figure 4.2**). In the dsDNA bound complex, a peptide corresponding to this feature (amino acids 345–358) exchanges less deuterium over time, than the same peptide in the unbound complex (**Figure 4.2B**). This indicates that the glutamine wedge is protected from exchange and is less dynamic when Cascade is bound to DNA. Similarly, a peptide covering the PAM sensing lysine finger (Cse1 amino acid 268) also exchanges less deuterium upon dsDNA binding. Results for the glutamine wedge and lysine finger are consistent with the structural models, however the glycine loop (Cse1 amino acids 157–161) showed no

increase in protection upon dsDNA binding, which was surprising since the crystal structure indicates that this feature interacts with the minor groove of the double-stranded PAM. This suggests, that in solution, the glycine loop may be more dynamic after DNA binding than is indicated by the crystal structure.

The Cse1 subunit is critical for binding DNA and the conformational state of this subunit has been implicated in recruitment and allosteric regulation of Cas3 *(124,143–145)*. In structures of Cascade without DNA or bound to dsDNA, the Cse1 subunit is tethered to Cas5 via a short loop (L1, amino acids 125–131) on Cse1 that is buried in a pore on Cas5 *(120)*. However, in the structure of Cascade bound to a complementary ssDNA target, the N-terminal domain (NTD) of Cse1 comes "unhitched" from the complex and L1 loop is disordered in the structure *(129)*. Collectively, these structures demonstrate that the NTD of Cse1 adopts conformationally distinct positions and recent FRET experiments by Xue et al. suggest that NTD of Cse1 is conformationally dynamic *(136)*. While the structures of Cascade without DNA suggest that Cse1 is in a "closed" conformation where L1 is buried in the Cas5 pore, our HDX data reveals that the L1 peptide of Cse1 (amino acids 118–129) is significantly more protected (less HD-exchange) in the dsDNA bound complex, as compared to the unbound complex (**Figure 4.2**). These data suggest that the Cse1 subunit is dynamic in the absence of DNA, sampling both the "open" and "closed" conformations.

Target recognition by Cascade relies on both specific- and non-sequence specific interactions with DNA. Recently, we showed that a lysine-rich helix (amino acid residues 137–144) on the fingers domain of two adjacent Cas7 subunits (i.e. Cas7.5 and Cas7.6)

creates a lysine-rich vise that makes non-sequence specific interactions with dsDNA and that these interactions are essential for target binding *(128)*. We predicted that dsDNA would result in less HD-exchange for peptides corresponding to the lysine-rich vise. To test this prediction, we examined a peptide covering the lysine-rich helices of Cas7 (amino acid residues 134–148). This peptide showed a small degree of protection upon dsDNA binding. However, there are six copies of the Cas7 subunit in the surveillance complex and only two of these subunits (Cas7.5 and Cas7.6) interact with the phosphate backbone of the incoming DNA *(128)*. The Cas7 peptide corresponding to the lysine-rich vise (residues 134–148) should have a different exchange rates depending on their position in the complex. Peptides from Cas7.5 and Cas7.6 should have less HD-exchange than the same peptides form Cas7.1–7.4, and this difference should result in a bimodal distribution of these peptides. In fact, we observe a bimodal behavior, which is consistent with a subset of the Cas7 subunits having increased protection upon DNA binding.

Path of the Displaced Strand

One of the outstanding questions about the mechanism of target binding is the location of the displaced DNA strand. Once Cascade binds to a dsDNA target, the complementary strand of the DNA target base pairs with the crRNA, while the non-complementary strand is displaced and forms an R-loop *(119,146)*. Structures of *E. coli* Cascade bound to dsDNA showed 10 to 12 nucleotides of the displaced DNA strand, but it remains unclear if this fragment reflects the path of the displaced strand when the complex binds to a full-length target duplex *(125)*. In the structure, the partial displaced strand is stabilized by interactions with residues on the Cse1 subunit of Cascade and we

expected that the remaining part of the displaced strand would interact with specific residues on Cas subunits as it traversed form the PAM proximal tail to the PAM distal head. These stabilizing interactions would shield specific amino acid residues from HD-exchange. To trace the path of the displaced strand, we used our HDX-MS data to identify peptides that show greater protection from exchange in the dsDNA bound form (**Figure 4.3A**).

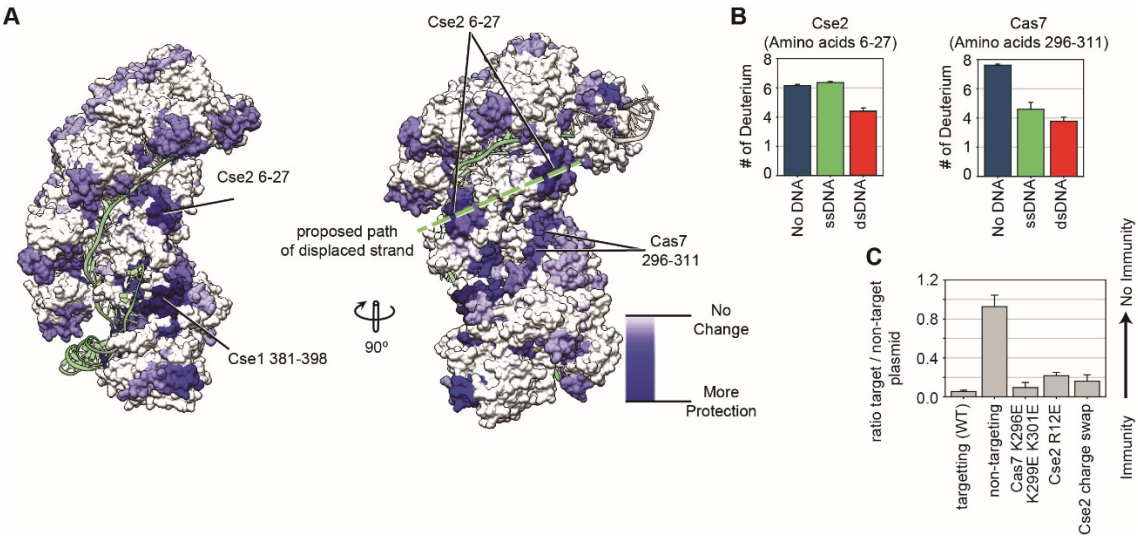

Figure 4.3 Using HDX to identify residues involved in binding the displaced DNA strand. (A) Differences in deuterium uptake between the unbound and dsDNA bound complex for individual peptides at the 180-minute time point are mapped onto the dsDNA bound structure . Only peptides that are more protected in the dsDNA bound complex (blue) are shown. The proposed path of the displaced strand is indicated (green). (B) Deuterium incorporation for peptides Cse2 6–27 and Cas7 296–311 at the 180-minute time point is displayed for Cascade without DNA (teal), ssDNA bound (green), and dsDNA bound complexes (red). The Cse2 peptide shows increased protection in the dsDNA bound complex, as compared to the unbound or the ssDNA bound complex ($p<0.01$). The Cas7 peptide shows the most protection in the dsDNA bound complex. Error bars represent the standard deviation of three replicates. (C) HDX-guided mutation of positively charged residues on Cse2 (R12E), Cse2 (R53D, K77D, K78D, R89D, R135D, R143D and K158D, named Cse2 charge swap), and Cas7 (K296E, K299E, K301E) were tested in a plasmid-curing assay to determine the biological impact of mutating these residues. The Cas7 mutant did not have a significant impact while the Cse2 mutants displayed a small, but significant defect. The single mutation of Cse2 R12E

had an equal impact as mutation of seven Cse2 residues. The error bars represent the standard error of the mean of three replicates.

Analysis of the tail of Cascade revealed a peptide in the C-terminal domain (CTD) of Cse1 (amino acid residues 381–398) with significant protection ($p<0.01$) in the dsDNA bound form of the complex compared to the unbound complex (**Figure 4.3A**) *(141)*. This is consistent with the path of the displaced strand in the dsDNA bound structure from Hayes *et al (125)*. However, to clarify the mechanism of protection at this location, we conducted an additional comparison by performing HDX-MS on Cascade bound to ssDNA. Comparison of the dsDNA and ssDNA bound complexes reveal regions of protection specifically resulting from the presence of the displaced strand. Importantly, this comparison showed equivalent levels of protection in the ssDNA bound and dsDNA bound form of Cascade, indicating that the increase in protection observed in the dsDNA bound structure is not necessarily due to direct interactions with the displaced strand, but likely an allosteric stabilization of the hydrogen-bonding network across this peptide. This result is consistent with a recent structure of *T. fusca* Cascade bound to a complete dsDNA target. In this structure, the R-loop is disordered in the region of the Cse1-CTD *(130)*.

Once the displaced strand leaves the four-helix bundle, it must travel up the complex where it is expected to form interactions with the belly or backbone subunits of the complex. The HDX data revealed two distinct areas of increased protection that may accommodate the displaced strand (**Figure 4.3A**). The upper route travels across the top of the Cse2 subunits and includes a peptide that shows strong protection (residues 6–27). The lower path of protection is formed by peptides covering residues 296–311 of the

Cas7 subunits. This lower route is part of the basic cleft that is formed by the Cse2 and Cas7 subunits and has previously been hypothesized to accommodate the displaced strand *(120,125,129)*. Interestingly, the Cse2 peptide corresponding to the "upper path" (amino acid residues 6–27) showed comparable levels of exchange in the unbound and ssDNA bound forms, while it was more protected from exchange in the dsDNA bound form ($p<0.001$), suggesting that this peptide may interact with the displaced strand (**Figure 4.3B**). Conversely, the Cas7 peptide corresponding to the "lower path" (amino acid residues 296–311) showed comparable levels of exchange in both the ssDNA and dsDNA bound states ($p>0.05$), suggesting that the enhanced stability measured by HDX upon dsDNA binding is not a result of stable interactions with the displaced strand (**Figure 4.3B**). To test the role of these residues in stabilizing the displaced strand we mutated arginine 12 on Cse2 to glutamic acid (Cse2 R12E) and three lysine residues on Cas7 to glutamic acids (Cas7 K296E, K299E, and K301E). We then tested the impact of these mutations using a plasmid-curing assay that measures the *in vivo* efficiency of Cascade/Cas3 to detect and degrade a target plasmid (**Figure 4.3C**) *(128,147)*. The Cse2 mutant (i.e. Cse2 R12E) displayed small, but reproducible defects, while the Cas7 mutant did not result in a measurable defect. We compared the HDX-guided mutants to a mutant we previously made based on electrostatic surface potential, where we mutated seven positively charged residues on Cse2 that line the basic cleft (Cse2 R53D, K77D, K78D, R89D, R135D, R143D and K158D, named Cse2 charge swap) (**Figure 4.3C**) *(120,129)*. The charge swap of seven Cse2 residues resulted in a small defect comparable to the

mutation of the single arginine 12 on Cse2. Collectively this suggests that precise coordination of the R-loop, may not be critical for immune system function.

Recruitment of the Nuclease-Helicase Cas3

DNA binding triggers a conformational change in Cascade that recruits Cas3 to the Cse1 subunit *(120,121,123,125,128,148)*. We hypothesized that dsDNA binding exposes a Cas3 docking site on Cascade and that these residues would display greater HD-exchange after Cascade binds dsDNA. To test this hypothesis, we compared HD-exchange before and after dsDNA binding and identified a large beta-hairpin, termed Loop 3 (L3, amino acids 285–296) on Cse1 that is more flexible after DNA binding (**Figure 4.4A**). The increased exchange in L3 coincides with movement of a short loop called L4 (amino acids 316–326). L4 is positioned next to L3 in the unbound and ssDNA bound Cascade structures, but moves 11 Å away from L4 upon dsDNA binding. To assess the functional importance of L3 and L4 we compared the Cse1 protein structure from *E. coli* to homologous protein structures from *Thermobifida fusca*, *Thermus thermophiles,* and *Acidimicrobium ferrooxidans (125,128,130,149–151)*. While there is no detectable amino acid sequence conservation, L3 and L4 are conserved structural features. Based on structural conservation and dynamic response to dsDNA binding, we hypothesized that L3 and L4 may participate in Cas3 recruitment. To test this hypothesis, we deleted L3 (L285-K296, ΔL3) and performed plasmid-curing assays. Deletion of L3 caused a plasmid-curing defect, indicating that L3 has an important functional role in DNA binding, Cas3 recruitment, or both (**Figure 4.4B**). To clarify the function of L3, we expressed and purified the Cse1ΔL3 deletion mutant of Cascade. Size exclusion

chromatography and SDS-PAGE gel electrophoresis confirmed that the complex assembles and purifies like wild type Cascade. However, Electrophoretic Mobility Shift Assays (EMSAs) revealed that the L3 deletion results in a DNA binding defect (K_D= 135.2 nM) compared to wild type Cascade (K_D= 1.5 nM) (**Figure 4.4C**), which indicates that defects in DNA binding are responsible for the immune defect observed in the plasmid-curing assays. To further investigate the function of L3, we focused on two conserved lysine residues (Cse1 K289 and K290) at the apex of L3. Alanine substitution of these two lysine residues (Cse1 K289A/K290A) resulted in a similar plasmid-curing defect as deletion of the entire 11 residue loop (**Figure 4.4B**). Like the ΔL3 mutant, the Cse1 K289A/K290A mutant expressed and purified similar to wild type. However, the Cse1 K289A/K290A mutation does not impair DNA binding (K_D= 1.8 nM) (**Figure 4.4C**). This suggested that the lysine residues might be involved in Cas3 recruitment or activation. To determine the role of the lysine residues, we performed an *in vitro* cleavage assay in which Cse1 K289A/K290A Cascade was pre-bound to dsDNA followed by the addition of Cas3. However, in contrast to our prediction, Cascade with the Cse1 K289A/K290A mutation results in a very modest DNA degradation defect (**Figure 4.4D**). We speculate that the discrepancy between the *in vivo* plasmid-curing assay (defect) and the *in vitro* DNA degradation assay (no defect) might be explained by high Cas3 concentrations (300 nM) and low of sensitivity in our DNA degradation assay. The mechanistic basis of this distinction requires further investigation, but the result illustrates the importance of performing both *in vivo* and *in vitro* experimentation.

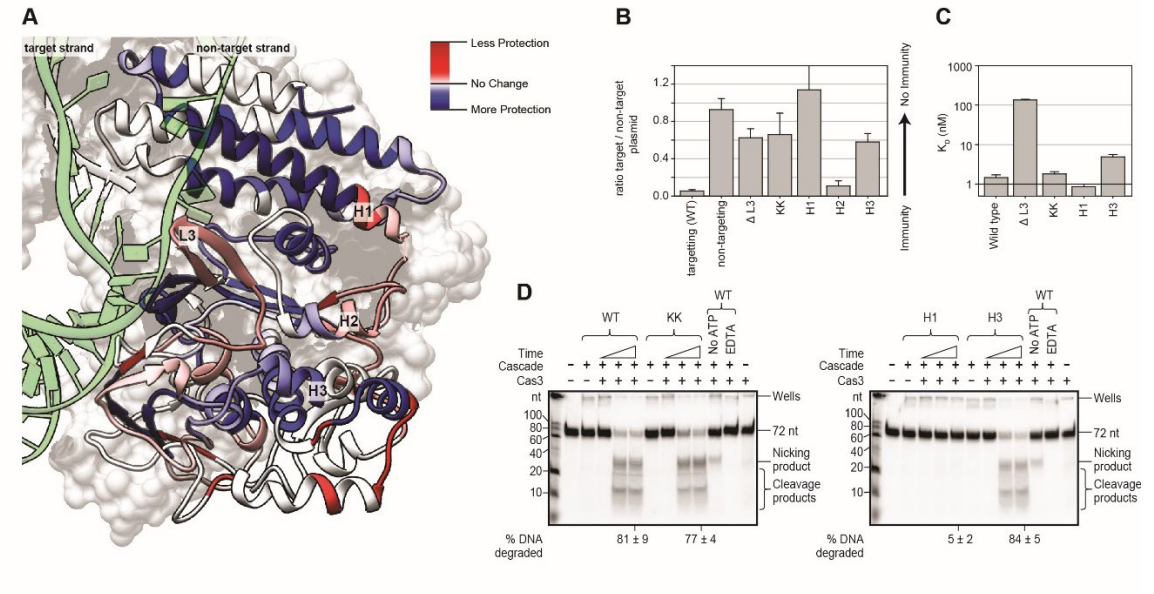

Figure 4.4. Predicting the Cas3 interaction site using HDX. (A) Ribbon representation of Cse1 colored according to the percentage difference in deuterium uptake of the dsDNA bound complex minus the unbound form of the Cascade complex at the 180-minute time point. Areas in red are more available for HDX after dsDNA binding. dsDNA is colored green. (B) Residues in three of the peptides (L3, H1 and H2) that exchange more deuterium after binding dsDNA and one peptide that exchanges less deuterium (H3) after binding dsDNA, were targeted for mutagenesis and tested using plasmid-curing assays. Deletion of Loop 3 (Cse1 ΔL285-K296, denoted ΔL3) or mutation of the lysine residues present at the tip of Loop 3 (Cse1 K289A/K290A, denoted KK) result in large immune system defects. Mutation of alpha helices H1 (Cse1 N379A/E380K), H2 (Cse1 N332A, Q333A, R335E) and H3 (Cse1 R194E/K197E) resulted in immune system defects for H1 and H3. Mutants showing defects were selected for further study. The error bars represent the standard error of the mean of three replicates. (C) Measurement of the binding affinity for 72 bp dsDNA containing a protospacer and interference PAM (3′-TAC-5′ or 3′-TTC-5′) using Electrophoretic Mobility Shift Assays (EMSAs). Only the deletion of Loop 3 resulted in a major binding defect. Error bars represent the standard deviation of three replicates. (D) Denaturing polyacrylamide gels showing time courses of Cascade/Cas3 mediated DNA degradation. 72 bp dsDNA containing P7 protospacer and 3′-TTC-5′ PAM was pre-bound to wild type Cascade or the Cascade mutants and incubated for various lengths of time with Cas3. Wild type Cascade facilitates destruction of the DNA by Cas3. The H1 mutant severely reduces the ability of Cas3 to degrade the DNA, indicating that this mutant blocks Cas3 recruitment or activation. The percentage of DNA degraded after 60 minutes is indicated with the error representing the standard deviation of three replicates.

In addition to L3, we identified other surface exposed peptides on Cse1 with changes in their HDX profile upon DNA binding. We made mutations in three distinct locations that cluster on an exposed surface of Cse1 that is anticipated to be involved in Cas3 recruitment (**Figure 4.4A**). Two alpha-helices (H1 and H2) that showed increased exchange upon dsDNA binding and one alpha helix (H3) with a slight decrease in HD-exchange were selected for mutational analysis. Polar residues with uncharged side chains were mutated to alanine residues, and residues with charged side chains were swapped for the opposite charge. The H1 mutation (Cse1 N379A/E380K) and the H3 mutation (Cse1 R194E/K197E) both result in pronounced defects in the plasmid-curing assay, while the H2 mutation (Cse1 N332A, Q333A and R335E) results in a modest immune system defect (**Figure 4.4B**). To clarify the functional importance of H1 and H3 we expressed and purified these two mutants. The expression levels and stoichiometry of the purified mutants were comparable to wild type Cascade. *In vitro* binding assays revealed that the H1 mutant binds dsDNA with wild type affinity (K_D=0.86 nM), while the H3 mutant has a small defect (K_D=4.87 nM) (**Figure 4.4C**). To test the function of these residues in Cascade-mediated recruitment of Cas3, we performed DNA cleavage assays. The H1 mutant binds DNA like wild type Cascade, but does not facilitate Cas3-mediated DNA degradation, while the H3 mutant, which has a modest binding defect, degrades DNA similar to wild type (**Figure 4.4D**). A high concentration of Cascade (300 nM) was used in the DNA cleavage assays to compensate for the small DNA binding defect observed. These data indicate that residues in H1 are critical for Cas3 recruitment

or activation, while the plasmid-curing defect of the H3 mutant is caused by the small DNA binding defect.

Discussion

Structures of Cascade have captured the complex in several distinct conformational states *(120,121,123,125,128,129,152)*. Collectively, these structural models reveal a series of conformational changes that coincide with DNA recognition and Cas3 recruitment. Here, we add an additional dimension to this work by performing HDX-MS on Cascade before and after target recognition. HDX-MS analysis reveals specific regions of the complex that have more or less HD-exchange after binding DNA. Mapping this data onto 3D-structures of Cascade highlights regions of functional importance that have been previously identified, as well as functionally important regions that do not stand out in static structures.

The lysine-rich vise of Cas7 and the PAM sensing domain of Cse1 are critical for DNA binding *(125,128)*. Slow deuteration of a peptide is indicative of low solvent exposure or extensive hydrogen bonding and we measure significant decreases in deuterium incorporation after dsDNA binding for peptides that cover the lysine-rich vise, and two the PAM sensing elements (i.e., lysine finger and glutamine wedge). However, the glycine-rich loop (amino acids 157–161), which has previously been shown to intercalate in the minor groove of the PAM duplex is not protected from HD-exchange in the dsDNA bound complex. These results suggest that the glycine-rich loop may be more dynamic after binding than indicated by the crystal structure.

While the mechanism of PAM sensing and directional hybridization of the crRNA guide to the DNA target is relatively well understood, the location of the displayed strand has not been determined for the *E. coli* Cascade complex. Based on the structure of *E. coli* Cascade bound to a partially duplexed DNA target, we expected the displaced strand to traverse over the C-terminal four-helix bundle of Cse1 and continue along a basic cleft formed by the Cse2 and Cas7 subunits (**Figure 4.3**) *(129)*. Indeed, peptides covering the C-terminal four-helix bundle of Cse1 (amino acids 381–398) incorporate significantly less deuterium after dsDNA binding. However, this same peptide is also protected (i.e. less deuterium incorporation) in the ssDNA bound complex, suggesting that the lack of deuterium uptake is due to changes in hydrogen bonding networks, rather than direct interactions with the displaced strand. This conclusion is consistent with the recent structure of *T. fusca* Cascade bound to a dsDNA target, which shows no density for the displaced DNA strand in this region *(130)*.

Following the displaced strand beyond the four-helix bundle we predicted that the remaining portion of the R-loop would be coordinated by residues that form a positively charged cleft between Cse2 and Cas7 *(129)*. To test the role of this cleft in DNA binding, we flipped the charge on seven positively charged residues on Cse2 (i.e. R53D, K77D, K78D, R89D, R135D, R143D and K158D) and three positively charged residues on Cas7 (i.e. K296E, K299E and K301E). Remarkably, neither of these mutants resulted in a pronounced immune defect (**Figure 4.3C**). Using the HD-exchange data, rather than focusing strictly on electrostatic surface potential, we identified a peptide on Cse2 (amino acids 6–27) that exchanges significantly ($p<0.001$) less deuterium upon binding of

dsDNA. Mutation of a single arginine (Cse2 R12) resulted in a small, but reproducible immune defect, which is consistent with a minor role for this peptide in stabilizing the displaced strand. In support of this conclusion, the recent *T. fusca* Cascade structure reveals an equivalently positioned arginine (R16) that is located ~4 Å from the phosphate backbone of the displaced strand *(130)*. While we initially expected to identify an extensive network of non-sequence specific, protein-mediated interactions involved in R-loop stabilization, our results and recent structures of the Cascade complex from *T. fusca* reveal a poorly ordered displaced strand that makes few direct interactions with the Cas proteins. Collectively this suggests that precise coordination of the R-loop at positions distal to the PAM, may not be critical for immune system function.

In addition to DNA binding, Cse1 also plays a critical role in Cas3 recruitment *(123,124,137)*. We hypothesized that regions of Cse1 that incorporate more deuterium after DNA binding, may be involved in Cas3 recruitment. To test this hypothesis, we designed HDX-guided mutations in L3, H1, and H2 regions of Cse1 (**Figure 4.4**). Mutations in L3 and H1 result in significant immune defects, while the H2 mutation has no measurable phenotype. We expected that the increase in HD-exchange would be indicative of regions involved in Cas3 recruitment; however, only the H1 mutant blocks Cas3-mediated DNA degradation, suggesting that this may be a docking site for Cas3.

Conformational dynamics are critical to the function of any molecular machine. Magnetic tweezers have been used to measure the dynamics of torque-dependent R-loop formation and DNA dissociation for Cascade *(146,153)*. These experiments reveal

conformational changes in the DNA target and recent FRET studies have been used to identify distinct conformational states of Cascade that are controlled by the sequence of the DNA target *(136,139,144)*. FRET has been used to show that binding to DNA targets containing specific PAM or protospacer mutations correspond with a rotation of the N-terminal domain in Cse1. This conformation is referred to as the "open" state, which does not effectively recruit nuclease-active Cas3, but recruits a nuclease-repressed form of Cas3, in the presence of Cas1 and Cas2 *(135)*. Collectively these studies suggest that the conformational state of Cascade is determined by the sequence of the DNA ligand, and that these conformations control the immune response by regulating the activity of Cas3 *(120,135,136)*.

While structure guided placement of FRET dyes provides a powerful means for measuring distances between two selected landmarks, this approach cannot provide a global understanding of the complex conformational dynamics of a large multi-subunit machine like Cascade. Here we demonstrate how the high-spatial resolution of FRET is complemented by the peptide-resolution of HDX-MS. In this paper, we validate the method by comparing our results to previous insights (e.g. PAM binding residues) and provide new information about the mechanism of DNA binding (e.g. coordination of the displaced strand), and Cas3 recruitment (e.g. role of H1). Future experiments performed using HDX-MS should focus on Cascade bound to DNA targets that have previously been shown to elicit interference or priming responses. These experiments will be critical to understanding how DNA-induced conformational states of Cascade control the downstream immune response.

Materials and Methods

Hydrogen-Deuterium Exchange Coupled with Mass Spectrometry (HDX-MS)

Three forms of the Cascade complex were tested using HDX-MS: unbound Cascade, Cascade bound to a ssDNA substrate, and Cascade bound to a dsDNA bound substrate. First Cascade was expressed and purified (see below) and pre-bound to ssDNA or dsDNA substrate, by incubating unbound Cascade with an excess of the appropriate oligonucleotide at 37 °C for 15 minutes. Excess oligonucleotides were removed from the ssDNA and dsDNA bound Cascade preparations by gel filtration as described below. Cascade (10mg/mL) was diluted 10-fold into reaction buffer (10mM sodium acetate, 50mM NaCl, pH 7.0) which was made in D_2O. Samples were removed at different time points and quenched to stop the exchange after 0, 0.16, 0.5, 2.5, 5.0, 10, 30 and 180 minutes. At each time point, 10 µL was withdrawn and placed into quenching/digestion solution at pH 2.5 containing 1% formic acid (FA, Sigma) and 0.2 mg/mL pepsin (Sigma) at 0°C. After one and half minutes of digestion, the tube was frozen in liquid N_2 and stored at −80°C until liquid chromatography-mass spectrometry (LC-MS) analysis. For Intact protein HDX-MS the reaction started in the same manner as the pepsin-based HDX-MS. Time points were collected at 0.5, 6, 11.5, 17, 22.5, 28, 33.5, 39, 44.5, 50, 55.5, 61, 66.5, 72, and 77.5 minutes. The intact protein HDX-MS time points were quenched in the same manner; however, the intact protein quench solution did not contain pepsin, and tubes were immediately frozen in liquid N_2 and stored at −80°C until liquid chromatography-mass spectrometry (LC-MS) analysis

LC-MS analysis of intact Cascade and Cascade peptides was completed on a 1290 UPLC series chromatography stack (Agilent Technologies) coupled directly to a 6538 UHD Accurate-Mass Q-TOF LC/MS mass spectrometer (Agilent Technologies). Before electrospray-time of flight (ESI-TOF) analysis, peptides were separated on a Reverse Phase (RP) column (Phenomenex Onyx Monolithic C18 column, 100 × 2 mm) at 1°C using a flow rate of 600 µL/min in the following conditions: 1min, 5% B; 1.0–9.0min, 5–45% B; 9.0–9.80 min, 95% B; 9.80–9.90 min, 5% B solvent A = 0.1% formic acid (FA, Sigma) in water (ThermoFisher) and solvent B = 0.1% FA in acetonitrile (ACN, ThermoFisher). Data was acquired at 2 Hz sec^{-1} over the scan range 50–1700 m/z in positive mode. Electrospray settings were: nebulizer set to 3.7 bar, drying gas 8.0 L/min, drying temperature 350°C, and capillary voltage 3.5 kV.

Data processing was carried out in Agilent MassHunter Qualitative Analysis version 6.0. Unique peptides were identified by using Peptide Analysis Worksheet version 2000.6.8.0 (PAWs, ProteoMetrics LLC.) with an accuracy of 10 ppm. Peptide sequences were also confirmed by MS-MS data acquisition, with a scan range of 50–1700 m/z (auto MS-MS)) with an isolation width of 4 m/z and an acquisition rate of 1 spectrum/s. Different collision energy voltage (20V, 25V, and 30V) and linear gradient voltages were applied in auto MS-MS mode. MS-MS data were analyzed and mapped to confirm the corresponding sequence by using Peptide Shaker version 0.41.1 paired with Search GUI version 1.30.1 (Compomics) *(107)*.

Deuterium incorporation in each peptide was calculated using the program HDExaminer (version 1.3.0 beta 6, Sierra Analytics Inc.) which measures the shift of the

centroid peak of the peptides collected throughout the time course in different samples. The shift of the centroid was then compared to a non-deuterated control (no exchange) and a fully deuterated sample (24-hour reaction) to determine the extent of exchange using methods previously reported by Guttman *et al. (140)*. Back-exchange during chromatography was estimated from a fully deuterated protein. The 24-hour time point was denoted as 100% ($m_{100\%}$) exchange. The non-deuterated control was denoted as 0% ($m_{0\%}$) exchange. Percentage deuteration for each peptide at different time points was calculated by following formula.

$$\%\, Deuteration = \frac{m_t - m_{0\%}}{m_{100\%} - m_{0\%}}$$

m_t = deuterium uptake at different times

Statistical analysis on deuterium uptake was performed using MEMHDX *(141)*. P-values less than 0.05 were considered statistically significant. For the uptake curves (**Figures 4.2 and 4.3**), peptides that were the most representative of the data were displayed. Uptake information for all peptides can be found in online in Supplementary Data Set 1.

Nested or sister peptides that shared the same N-terminal or C-terminal residue and varied in length were manually resolved to determine the contributions of the different amides to the overall uptake of the peptides. The overlapping regions of the peptides were assumed to take up the same amount of deuterium and the non-overlapping regions were assumed to be responsible for the rest of the uptake in the longer peptide. For example, if the peptide covering residues 1–4 took up two deuterium and the peptide covering residues 1–6 took up three deuterium, it would be assumed that residues 5 and 6

together took up, on average, one deuterium. The uptake percentage of the resolved peptides of the unbound complex were then subtracted from the dsDNA bound complex and mapped onto the structural model of dsDNA bound Cascade (5H9F) using Chimera version 1.10.2 (UCSF Resource for Biocomputing, Visualization, and Informatics), to create the difference maps shown in Figures 1-4 *(154)*.

Cascade Expression and Purification

Cascade and Cascade mutants were expressed and purified using previously described methods *(128)*. Briefly, *cas* genes and CRISPR RNAs were coexpressed in *E. coli* Bl21 (DE3) cells with Cse2 fused to an N-terminal Strep tag. Cells were grown in LB media at 37°C under antibiotic selection and induced at an OD_{600nm} of 0.5 using 0.2 mM isopropyl-D-1-thiogalactopyranoside (IPTG). Cells were cultured overnight at 20 °C, pelleted by centrifugation and suspended in lysis buffer (100 mM Tris–HCl pH 8.0, 150 mM NaCl, 1 mM Ethylenediaminetetraaceticacid (EDTA), 1 mM tris (2-carboxyethyl) phosphine (TCEP) and 5% glycerol), and frozen at −80 °C. Cells were lysed by sonication and lysates were clarified by centrifugation. Cascade self-assembles in vivo and the complex was affinity purified on StrepTrap HP resin (GE) using N-terminal Strep-II tags on Cse2. Elution of Cascade was performed using the lysis buffer supplemented with 2.5 mM desthiobiotin. The Strep-II tag was removed with HRV-3C protease. Cascade was concentrated prior to gel filtration chromatography using 10/300 Superose6, 16/60 Superose6 or 26/60 Superose6 columns (GE Healthcare) equilibrated with 50 mM Tris–HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM TCEP and 5% glycerol and stored at −80 °C until use. Stoichiometry of the different Cascade mutants

was analyzed by SDS-PAGE gel and the presence of crRNA was checked by phenol/chloroform extraction of the protein followed by denaturing polyacrylamide gel.

Cas3 Expression and Purification

Cas3 was expressed and purified using previously described methods *(122,123)*. Briefly, the Cas3 gene fused to an N-terminal His_6-MBP (maltose binding protein) tag was expressed from pHMGWA plasmid in *E. coli* BL21 (DE3) cells. Cells were grown in LB media at 20°C under antibiotic selection and expression was induced at an $OD_{600\ nm}$ of 0.3 using 0.2 mM IPTG. Cells were pelleted by centrifugation and suspended in lysis buffer (20 mM TRIS-HCl pH 8.0, 200 mM NaCl, 1 mM TCEP and 10 % glycerol). Cells were lysed by sonication and lysates were clarified by centrifugation. Cas3 was affinity purified using Ni-NTA superflow resin (Qiagen), washed with lysis buffer supplemented with 5 mM Imidazole followed by a 1 M Nacl wash. Cas3 was eluted with lysis buffer supplemented with 250 mM Imidazole. Cas3 was concentrated prior to gel filtration chromatography using 16/60 Superoses6 column equilibrated with 20 mM TRIS-HCl pH 8.0, 200 mM NaCl, 1 mM TCEP and 5 % glycerol and stored at −80°C until use. Protein purity was checked by SDS-PAGE gel.

Oligoduplex Preparation and DNA Labeling

Oligonucleotides were 5-end labeled with [γ-^{32}P] ATP (PerkinElmer) using T4 polynucleotide kinase (NEB). Labeled oligonucleotides were purified by phenol/chloroform extraction followed by MicroSpin G-25 columns (GE Healthcare). Oligonucleotides were gel purified on a denaturing polyacrylamide gel and recovered. dsDNA was prepared by mixing labeled oligonucleotides with a 5-fold molar excess of

the complementary oligonucleotide in hybridization buffer (20 mM HEPES pH 7.5, 100 mM KCl and 5 % glycerol). The mixture was heated to 95°C for 5 minutes and allowed to cool to room temperature. Duplexed oligonucleotides were purified on a native polyacrylamide gel and recovered in hybridization buffer. Duplexed oligonucleotides were used for EMSA and Cascade/Cas3 mediated DNA degradation assays.

Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays were carried out as described previously *(128)*. The reported K_Ds are the average from three independent experiments and error bars represent the standard deviation. Oligonucleotides used for EMSAs are listed in Supplemental table 1 which can be found online. The non-target DNA strand was labeled.

Plasmid-Curing Assays

Plasmid-curing assays were carried out as described previously *(128)*. Briefly, competent *E. coli* cells containing plasmids encoding Cascade, Cas3, and a crRNA targeting the pUC-P7-TAC plasmid were transformed with an equal molar mixture of pUC19 (non-target) and pUC-p7-TAC (target) plasmids. Expression of Cascade, Cas3, and the crRNA was induced using 0.2 mM isopropyl-D-1-thiogalactopyranoside (IPTG) and cells were plated on LB-agar plates. Forty individual colonies were screened using PCR designed to discriminate target from non-target plasmids based on size. Three replicates were carried out for each experiment and the average value of the ratio of target to non-target plasmid was calculated and plotted in the graphs (**Figures 4.3 and 4.4**). An active immune system will destroy the target plasmid and result in a value close

to 0, while an inactive immune system will preserve the original 1:1 ratio of target to non-target plasmid and result in a value close to 1. The error bars represent standard error of the mean calculated from three independent experiments.

Cascade/Cas3 Mediated DNA Degradation

Duplexed oligonucleotides were [γ-^{32}P] ATP labeled on the 5′end of the non-target DNA strand and incubated with 300 nM Cascade or Cascade mutants for 15 minutes at 37°C in reaction buffer (20 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 100 μM $NiSO_4$, 100 μM $CoCl_2$ and 2 mM TCEP) supplemented with 2 mM ATP (Lanes marked No ATP did not receive ATP, lanes marked EDTA contained 17 mM EDTA, **Figure 4.4D**). Reactions were incubated on ice for 5 minutes prior to addition of Cas3 to a final concentration of 300 nM. After the addition of Cas3 reactions were incubated for 60 minutes at 37 °C except for the samples that were part of the time course (triangle in **Figure 4.4D**, incubated for 0, 15 and 60 minutes). Once the reactions were complete the samples were quenched by the addition of EDTA and Sodium Dodecyl Sulphate (SDS) to a final concentration of 8 mM EDTA and 1 % SDS. Samples were mixed with 2x Formamide loading buffer, heated to 95°C C for 5 minutes and run on a pre-run 7 M Urea, 15 % polyacrylamide (29:1) gel in Tris-Taurine-EDTA (TTE) buffer. Gels were dried, exposed to phosphor storage screens and scanned with a Typhoon phosphorimaginer. Degraded and undegraded DNA fractions were quantified using GelQuant.NET software and used to calculate the fraction of degraded DNA compared to the total DNA. Reported degradation percentages are the average of three independent experiments with the error representing the standard deviation.

References

(118) Brouns, S. J. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J. H.; Snijders, A. P. L.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der OOst, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **2013**, *321* (5891), 960–965. https://doi.org/10.1126/science.1159689.

(119) Jore, M. M.; Lundgren, M.; Van Duijn, E.; Bultema, J. B.; Westra, E. R.; Waghmare, S. P.; Wiedenheft, B.; Pul, Ü.; Wurm, R.; Wagner, R.; Beijer, M. R.; Barendregt, A.; Zhou, K.; Snijders, A. P. L.; Dickman, M. J.; Doudna, J. A.; Boekema, E. J.; Heck, A. J. R.; Van Der Oost, J.; Brouns, S. J. J. Structural Basis for CRISPR RNA-Guided DNA Recognition by Cascade. *Nat. Struct. Mol. Biol.* **2011**, *18* (5), 529–536. https://doi.org/10.1038/nsmb.2019.

(120) Jackson, R. N.; Golden, S. M.; Erp, P. B. G. Van; Carter, J.; Westra, E. R.; Brouns, S. J. J.; Oost, J. Van Der; Terwilliger, T. C.; Read, R. J.; Wiedenheft, B. Crystal Structure of the CRISPR RNA– Guided Surveillance Complex from Escherichia Coli. *Science (80-.).* **2014**, *345* (6203), 1473–1479. https://doi.org/10.1126/science.1256328.

(121) Zhao, H.; Sheng, G.; Wang, J.; Wang, M.; Bunkoczi, G.; Gong, W.; Wei, Z.; Wang, Y. Crystal Structure of the RNA-Guided Immune Surveillance Cascade Complex in Escherichia Coli. *Nature* **2014**, *515* (7525), 147–150. https://doi.org/10.1038/nature13733.

(122) Mulepati, S.; Bailey, S. In Vitro Reconstitution of an Escherichia Coli RNA-Guided Immune System Reveals Unidirectional, ATP-Dependent Degradation of DNA Target. *J. Biol. Chem.* **2013**, *288* (31), 22184–22192. https://doi.org/10.1074/jbc.M113.472233.

(123) Hochstrasser, M. L.; Taylor, D. W.; Bhat, P.; Guegler, C. K.; Sternberg, S. H.; Nogales, E.; Doudna, J. A. CasA Mediates Cas3-Catalyzed Target Degradation during CRISPR RNA-Guided Interference. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (18), 6618–6623. https://doi.org/10.1073/pnas.1405079111.

(124) Westra, E. R.; van Erp, P. B. G.; Künne, T.; Wong, S. P.; Staals, R. H. J.; Seegers, C. L. C.; Bollen, S.; Jore, M. M.; Semenova, E.; Severinov, K.; de Vos, W. M.; Dame, R. T.; de Vries, R.; Brouns, S. J. J.; van der Oost, J. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* **2012**, *46* (5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018.

(125) Hayes, R. P.; Xiao, Y.; Ding, F.; Van Erp, P. B. G.; Rajashankar, K.; Bailey, S.; Wiedenheft, B.; Ke, A. Structural Basis for Promiscuous PAM Recognition in Type I-E Cascade from E. Coli. *Nature* **2016**, *530* (7591), 499–503.

https://doi.org/10.1038/nature16995.

(126) Semenova, E.; Jore, M. M.; Datsenko, K. A.; Semenova, A.; Westra, E. R.; Wanner, B.; van der Oost, J.; Brouns, S. J. J.; Severinov, K. Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence. *Proc. Natl. Acad. Sci.* **2011**, *108* (25), 10098–10103. https://doi.org/10.1073/pnas.1104144108.

(127) Jung, C.; Hawkins, J. A.; Jones, S. K.; Xiao, Y.; Rybarski, J. R.; Dillard, K. E.; Hussmann, J.; Saifuddin, F. A.; Savran, C. A.; Ellington, A. D.; Ke, A.; Press, W. H.; Finkelstein, I. J. Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **2017**, *170* (1), 35-47.e13. https://doi.org/10.1016/j.cell.2017.05.044.

(128) van Erp, P. B. G.; Jackson, R. N.; Carter, J.; Golden, S. M.; Bailey, S.; Wiedenheft, B. Mechanism of CRISPR-RNA Guided Recognition of DNA Targets in Escherichia Coli. *Nucleic Acids Res.* **2015**, *43* (17), 8381–8391. https://doi.org/10.1093/nar/gkv793.

(129) Mulepati, S.; Héroux, A.; Bailey, S. Structural Biology. Crystal Structure of a CRISPR RNA-Guided Surveillance Complex Bound to a SsDNA Target. *Science* **2014**, *345* (6203), 1479–1484. https://doi.org/10.1126/science.1256996.

(130) Xiao, Y.; Luo, M.; Hayes, R. P.; Kim, J.; Ng, S.; Ding, F.; Liao, M.; Ke, A. Structure Basis for Directional R-Loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **2017**, *170* (1), 48-60.e11. https://doi.org/10.1016/j.cell.2017.06.012.

(131) Fineran, P. C.; Gerritzen, M. J. H.; Suárez-Diez, M.; Künne, T.; Boekhorst, J.; van Hijum, S. A. F. T.; Staals, R. H. J.; Brouns, S. J. J. Degenerate Target Sites Mediate Rapid Primed CRISPR Adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (16), E1629–E1638. https://doi.org/10.1073/pnas.1400071111.

(132) Westra, E. R.; Semenova, E.; Datsenko, K. A.; Jackson, R. N.; Wiedenheft, B.; Severinov, K.; Brouns, S. J. J. Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genet.* **2013**, *9* (9), e1003742–e1003742. https://doi.org/10.1371/journal.pgen.1003742.

(133) Datsenko, K. A.; Pougach, K.; Tikhonov, A.; Wanner, B. L.; Severinov, K.; Semenova, E. Molecular Memory of Prior Infections Activates the CRISPR/Cas Adaptive Bacterial Immunity System. *Nat. Commun.* **2012**, *3*, 945. https://doi.org/10.1038/ncomms1937.

(134) Richter, C.; Dy, R. L.; McKenzie, R. E.; Watson, B. N. J.; Taylor, C.; Chang, J. T.; McNeil, M. B.; Staals, R. H. J.; Fineran, P. C. Priming in the Type I-F CRISPR-

Cas System Triggers Strand-Independent Spacer Acquisition, Bi-Directionally from the Primed Protospacer. *Nucleic Acids Res.* **2014**, *42* (13), 8516–8526. https://doi.org/10.1093/nar/gku527.

(135) Redding, S.; Sternberg, S. H.; Marshall, M.; Gibb, B.; Bhat, P.; Guegler, C. K.; Wiedenheft, B.; Doudna, J. A.; Greene, E. C. Surveillance and Processing of Foreign DNA by the Escherichia Coli CRISPR-Cas System. *Cell* **2015**, *163* (4), 854–865. https://doi.org/10.1016/j.cell.2015.10.003.

(136) Xue, C.; Whitis, N. R.; Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell* **2016**, *64* (4), 826–834. https://doi.org/10.1016/j.molcel.2016.09.033.

(137) Xue, C.; Seetharam, A. S.; Musharova, O.; Severinov, K.; Brouns, S. J. J.; Severin, A. J.; Sashital, D. G. CRISPR Interference and Priming Varies with Individual Spacer Sequences. *Nucleic Acids Res.* **2015**, *43* (22), 10831–10847. https://doi.org/10.1093/nar/gkv1259.

(138) Swarts, D. C.; Mosterd, C.; van Passel, M. W. J.; Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **2012**, *7* (4), e35888–e35888. https://doi.org/10.1371/journal.pone.0035888.

(139) Blosser, T. R.; Loeff, L.; Westra, E. R.; Vlot, M.; Künne, T.; Sobota, M.; Dekker, C.; Brouns, S. J. J.; Joo, C. Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell* **2015**, *58* (1), 60–70. https://doi.org/10.1016/j.molcel.2015.01.028.

(140) Guttman, M.; Weis, D. D.; Engen, J. R.; Lee, K. K. Analysis of Overlapped and Noisy Hydrogen/Deuterium Exchange Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (12), 1906–1912. https://doi.org/10.1007/s13361-013-0727-5.

(141) Hourdel, V.; Volant, S.; O'Brien, D. P.; Chenal, A.; Chamot-Rooke, J.; Dillies, M.-A.; Brier, S. MEMHDX: An Interactive Tool to Expedite the Statistical Validation and Visualization of Large HDX-MS Datasets. *Bioinformatics* **2016**, *32* (22), 3413–3419. https://doi.org/10.1093/bioinformatics/btw420.

(142) Engen, J. R. Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS. *Anal. Chem.* **2009**, *81* (19), 7870–7875. https://doi.org/10.1021/ac901154s.Analysis.

(143) Xue, C.; Whitis, N. R.; Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell* **2016**, *64* (4), 826–834. https://doi.org/10.1016/j.molcel.2016.09.033.

(144) Jackson, R. N.; van Erp, P. B.; Sternberg, S. H.; Wiedenheft, B. Conformational Regulation of CRISPR-Associated Nucleases. *Curr. Opin. Microbiol.* **2017**, *37*,

110–119. https://doi.org/10.1016/j.mib.2017.05.010.

(145) Sashital, D. G.; Jinek, M.; Doudna, J. A. An RNA-Induced Conformational Change Required for CRISPR RNA Cleavage by the Endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **2011**, *18* (6), 680–687. https://doi.org/10.1038/nsmb.2043.

(146) Szczelkun, M. D.; Tikhomirova, M. S.; Sinkunas, T.; Gasiunas, G.; Karvelis, T.; Pschera, P.; Siksnys, V.; Seidel, R. Direct Observation of R-Loop Formation by Single RNA-Guided Cas9 and Cascade Effector Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (27), 9798–9803. https://doi.org/10.1073/pnas.1402597111.

(147) Almendros, C.; Guzmán, N. M.; Díez-Villaseñor, C.; García-Martínez, J.; Mojica, F. J. M. Target Motifs Affecting Natural Immunity by a Constitutive CRISPR-Cas System in Escherichia Coli. *PLoS One* **2012**, *7* (11), e50797–e50797. https://doi.org/10.1371/journal.pone.0050797.

(148) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. RNA-Guided Genetic Silencing Systems in Bacteria and Archaea. *Nature* **2012**, *482* (7385), 331–338. https://doi.org/10.1038/nature10886.

(149) Mulepati, S.; Orr, A.; Bailey, S. Crystal Structure of the Largest Subunit of a Bacterial RNA-Guided Immune Complex and Its Role in DNA Target Binding. *J. Biol. Chem.* **2012**, *287* (27), 22445–22449. https://doi.org/10.1074/jbc.C112.379503.

(150) Sashital, D. G.; Wiedenheft, B.; Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **2012**, *46* (5), 606–615. https://doi.org/10.1016/j.molcel.2012.03.020.

(151) Tay, M.; Liu, S.; Yuan, Y. A. Crystal Structure of Thermobifida Fusca Cse1 Reveals Target DNA Binding Site. *Protein Sci.* **2015**, *24* (2), 236–245. https://doi.org/10.1002/pro.2609.

(152) Wiedenheft, B.; Lander, G. C.; Zhou, K.; Jore, M. M.; Brouns, S. J. J.; Van Der Oost, J.; Doudna, J. A.; Nogales, E. Structures of the RNA-Guided Surveillance Complex from a Bacterial Immune System. *Nature* **2011**, *477* (7365), 486–489. https://doi.org/10.1038/nature10402.

(153) Rutkauskas, M.; Sinkunas, T.; Songailiene, I.; Tikhomirova, M. S.; Siksnys, V.; Seidel, R. Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep.* **2015**, *10* (9), 1534–1543. https://doi.org/10.1016/j.celrep.2015.01.067.

(154) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

https://doi.org/10.1002/jcc.20084.

CHAPTER FIVE

THERMODYNAMICS OF CRISPR-ANTI-CRISPR INTERACTIONS PROVIDES MECHANISTIC INSIGHT INTO INHIBITION

Contribution of Authors and Co-Authors

Manuscript in Chapter 5

Author: Angela Patterson

Contributions: Designed and performed HDX-MS and DSF experiments, wrote the manuscript

Co-Author: Aidan White

Contributions: Aided with HDX-MS and DSF experiments, edited the manuscript

Co-Author: Elizabeth Waymire

Contributions: Aided with HDX-MS and DSF experiments, edited the manuscript

Co-Author: Sophie Fleck

Contributions: Aided with HDX-MS experiments, edited the manuscript

Co-Author: Sarah M. Golden

Contributions: Purified Csy, AcrIF9 bound, and ssDNA bound Csy complexes

Co-Author: Royce Wilkinson

Contributions: Purified AcrIF2 bound and unbound Csy complexes, edited the manuscript

Co-Author: Blake Wiedenheft

Contributions: Supervised purification of protein complexes, edited the manuscript

Co-Author: Brian Bothner

Contributions: Supervised HDX-MS and DSF experiments, wrote the manuscript

Manuscript Information

Angela Patterson, Aidan White, Elizabeth Waymire, Sophie Fleck, Sarah Golden, Royce Wilkinson, Blake Wiedenheft, Brian Bothner

To be determined

Status of Manuscript:
__x__ Prepared for submission to a peer-reviewed journal
_____ Officially submitted to a peer-reviewed journal
_____ Accepted by a peer-reviewed journal
_____ Published in a peer-reviewed journal

Abstract

CRISPR RNA-guided detection and degradation of foreign DNA is a dynamic process. This process is interrupted during viral infection through the binding of small viral proteins called anti-CRISPRs. While the structure of many anti-CRISPRs bound to their target CRISPR system have been solved, questions remain centering around the mechanism of how these small viral proteins disrupt the interference step of their associated CRISPR systems. Here we use hydrogen deuterium exchange-mass spectrometry (HDX-MS) and differential scanning fluorimetry (DSF) to measure changes to the conformational landscape of the type IF CRISPR RNA guided surveillance complex (Csy) upon binding of two different anti-CRISPR proteins (AcrIF9 and AcrIF2). While both anti-CRISPRs block sequence specific detection of DNA, binding may have further implications in Csy function. To determine how these anti-CRISPR proteins effect the Csy, a thermodynamic analysis of the complex either bound or unbound were probed. HDX-MS was used to track changes in protein dynamics as a proxy for understanding conformational entropy, while DSF was used to determine changes in enthalpy. The strength of the interaction between either AcrIF2 or AcrIF9 and the Csy complex is the driving force behind the ability of the Acr to inhibit the binding of dsDNA to Csy. The results from our experiments demonstrate that AcrIF2 binding relies on enthalpic stabilization, whereas AcrIF9 uses entropic stabilization to bind to Csy. This shows that CRISPR systems are an opportune model to dissect the nature of protein-protein interaction and the thermodynamic basis of molecular recognition by monitoring the

ability of viruses to take advantage of evolutionary space to arrive at different solutions to the same problem.

Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) and associated *cas* genes are essential components of adaptive immune systems that protect bacteria and archaea from foreign genetic elements *(155)*. In Class 1 CRISPR systems, the crRNA array is transcribed and processed by the CRISPR associated (Cas) protein, Cas6, into CRISPR RNA (crRNA) *(118,156)*. Once processed, multiple Cas proteins assemble around the crRNA forming the surveillance complex. This complex surveys the cell, looking for foreign DNA that matches the crRNA guide *(118)*. In the Type IF system found in *Pseudomonas aeruginosa*, the CRISPR surveillance complex is called Csy. It consists of crRNA surrounded by a Cas6f protein at the head of the complex, six Cas7 subunits which form the backbone in the middle of the complex, and a Cas5 and Cas8f subunit at the tail of the complex (**Figure 5.1**) *(157)*. Upon recognition of target dsDNA by the surveillance complex, Cas3, a helicase/endonuclease is recruited to degrade the foreign DNA, protecting the cell from infection *(122,124,158)*.

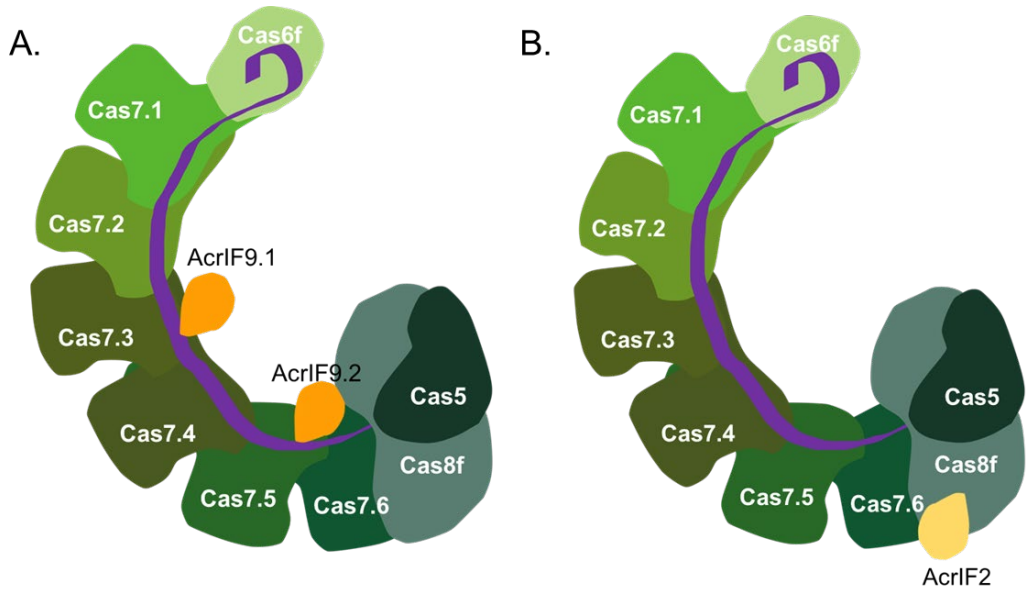

Figure 5.1. Cartoon depictions of the Csy complex either bound to AcrIF9 (A) or AcrIF2 (B). Two AcrIF9 proteins bind to the Csy complex along the Cas7 backbone while one AcrIF2 protein binds at the tail of the complex interacting with Cas7.6 and Cas8.

To escape the protection afforded by CRISPR immune systems, viruses have evolved small (<150 amino acid residues) proteins which interact with the Cas proteins to prevent DNA degradation *(159)*. These proteins, called anti-CRISPRs (Acr), work through a variety of mechanisms including sterically blocking DNA from binding to the complex, acting as a DNA mimic, inhibiting the helicase/endonuclease Cas3, or covalently modifying key Cas proteins *(160–163)*. While similar in size, they show no homology to each other or other known proteins *(164)*. The anti-CRISPR protein AcrIF9 binds Cas7 subunits in the middle and at the tail of the Csy complex (**Figure 5.1**) *(165,166)*. AcrIF9 interacts with dsDNA recognition residues, thus preventing specific binding of dsDNA to the complex *(166)*. Another anti-CRISPR of the Type IF system, AcrIF2, acts as a DNA mimic. AcrIF2 has a structure and electrostatic surface potential

reminiscent of B-form DNA. Acidic residues on the anti-CRISPR interact with lysine rich regions on Cas7.6 and Cas8f of the Csy complex*(162)* (**Figure 5.1**). This area is critical for initial DNA binding. Similar to AcrIF9, the structure and binding sites of AcrIF2 are known. However, the effects of these anti-CRISPRs on Csy complex dynamics and stability are unknown.

Recent structural studies have answered questions about where AcrIF2 and AcrIF9 bind to Csy *(162,166)*, while raising others. Both of the Acr bound complexes are similar in structure, however, there appears to be slight conformational changes which occur distal to the Acr binding site. Though these conformational changes are slight, the presence of long-distant changes indicates that allostery may be involved in Acr function. This idea is supported by our previous work on the Type IE CRISPR interference complex, Cascade, which demonstrated a role for allostery in nuclease recruitment *(109)*. Does the allosteric network which may be present in Csy, play a role in Acr binding? Many proteins exist as an ensemble of conformations in the native state. The ensemble can be described by an energy landscape on which highly populated conformations reside in deep wells while conformations which contribute less to the ensemble are in shallow wells *(4)*. The binding of ligands to a protein or protein complex alters the conformational landscape favoring the ligand bound conformation *(10)*. By comparing the structure and dynamics of ligand bound and unbound forms, functional properties of the conformational ensemble, as well as the basis for allostery can be elucidated.

Hydrogen deuterium exchange coupled to mass spectrometry (HDX-MS) is a powerful tool for studying protein stability and dynamics *(21,167)*. HDX-MS takes

advantage of the chemical tendency for amide hydrogens on the protein backbone to exchange with solvent *(168)*. When a protein is incubated in deuterated buffer, this exchange event can be monitored using mass spectrometry to track the increase in mass due to the exchange of a deuteron for a proton. Exchange can only occur if the amide hydrogen is solvent exposed and is not participating in a hydrogen bond *(20)*. Initially protected amide hydrogens can exchange through local or global conformational change in which hydrogen bonds temporarily break *(21)*. By comparing the uptake of deuterium between conditions, changes in solvent accessibility, secondary structure stability, and protein dynamics can be inferred. These changes can lead to insights into anti-CRISPR mechanisms and the potential role of allostery in blocking dsDNA binding.

Differential scanning fluorimetry (DSF) analyzes protein thermal stability. In DSF, proteins are unfolded by slowly raising the temperature (for example 1°C/minute from 25 to 90°C) exposing hydrophobic regions. Using a dye that increases fluorescence in nonpolar environments, thermal stability can be monitored by tracking fluorescence intensity *(110,111)*. Intensity *vs* temperature plots are used to assign a melting temperature which can be used to compare thermal stability.

By measuring changes in local and global protein backbone hydrogen exchange along with thermal stability, we determined that AcrIF9 and AcrIF2 alter Csy in very different ways. Our in-solution analysis using hydrogen-deuterium exchange mass spectrometry (HDX-MS) and differential scanning fluorimetry (DSF) reveal thermodynamically distinct outcomes of Acr binding. While the global conformation of the Acr bound complex does not change dramatically from that of unbound Csy, the

stability and dynamics are significantly altered. Here in, we present a model for Csy inhibition by AcrIF9 and AcrIF2 in which allosteric effects are driven by entropic or enthalpic modulation, respectively. Our model is consistent with recent ensemble theories of allostery and highlights that these two Acrs function through fundamentally different mechanisms.

Methods

Purification of Csy Complex

The Csy complex was purified as described previously *(169)*. Csy genes and a synthetic CRISPR were co-expressed on separate vectors (Addgene IDs: 89232 and 89244) in *Escherichia coli* BL21 (DE3) cells (NEB). Expression was induced with 0.5 mM isopropyl-D-1-thiogalactopyranoside (IPTG) at an optical density (OD_{600} nm) of 0.5. Cells were incubated overnight at 16 °C, then pelleted by centrifugation (3000 × g for 10 min at 4 °C) and re-suspended in lysis buffer containing 50 mM Tris at pH 7.5 and 500 mM sodium chloride. Pellets were sonicated on ice and the lysate was clarified by centrifugation at 10,000 × g for 25 min at 4 °C. The Csy complex self-assembles in vivo and the His-tagged complex (N-terminal 6-histidine affinity tags on Cas7f) was affinity purified using Ni-NTA resin (QIAGEN). The resin was washed with five column volumes of lysis buffer before elution with 50 mM Tris (pH = 7.5), 300 mM NaCl, 250 mM imidazole, 5% glycerol. Protein was then concentrated (Corning Spin-X concentrators) at 4 °C prior to size-exclusion chromatography (Superdex 200, GE Healthcare) in 20 mM HEPES, pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM TCEP.

ssDNA Bound Csy Preparation

Oligos for the target strand (GCT GTA CGT CAC TAT CGA AGC AAT A**CA GGT AGA CGC GGA CAT CAA GCC CGC CGT GAA** GGT GCA GCT TCT CTA CAG AGT GC) were annealed to the Csy complex.

Purification of AcrIF9 Bound Csy

AcrIF9 bound Csy was purified as described previously *(169)*. Briefly, Csy genes (Addgene ID: 89232), a synthetic CRISPR (Addgene ID: 89244), and AcrIF9 (GenBank: EEG86164.1 cloned into pCDF-1b) were co-expressed on separate vectors, as described above. The Csy–AcrIF9 complex was purified using methods similar to those described for the Csy complex, except size exclusion was performed using a Superdex 200 26/600 column (GE Healthcare), equilibrated in 20 mM HEPES, pH 7.5, 100 mM NaCl, 5% glycerol, and 1 mM TCEP.

Purification of AcrIF2 Bound Csy

AcrIF2 bound Csy was purified as described previously *(170)*. Briefly, Gene 30 from phage D3112 (AcrIF2) was cloned in a p15TV-L vector with N-terminal His6tags and overexpressed in E. coli BL21 DE3 cells as previously described (Bondy-Denomy et al., 2015). Briefly, E. coli BL21 DE3 cells were grown to an OD600 nm of 0.4 and then induced with IPTG at 16ºC for 16h. Cells were collected by centrifugation and suspended in a lysis buffer containing 50 mM Tris, pH 7.5, 500 mM NaCl). The cells were lysed by sonication and the lysate was centrifuged at 10,000 × g for 25 min to remove cell debris. The supernatant was injected into a Ni-NTA column, washed with lysis buffer supplemented with 20 mM imidazole and eluted from the column using a using a linear

gradient to 300 mM imidazole in the wash buffer. Fractions were collected, and concentrated (Corning Spin-X concentrators) at 4ºC before cleavage of the purification tag with TEV protease during dialysis against SEC buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 1 mM TCEP, 5% glycerol). The cleaved AcrIF2 was separated from the His-tag and uncleaved product by passing through the Ni-NTA column again. A 4-fold molar excess of the cleaved AcrIF2 was mixed with purified Csy complex and incubated at 37 ºC for 15 min and free Acr protein was separated from the Csy-AcfIF2 complex using a Superdex 200 size-exclusion column (GE Healthcare) in SEC buffer.

Intact HDX-MS

Concentrated protein stocks, Csy (9.7 μM), Csy with ssDNA (15.7μM), Csy with AcrIF9 (13.7 μM), and Csy with AcrIF2 (35.7 μM) were diluted 1:10 into deuterated 50 mM HEPES 150 mM KCl, pH 7.5 or into non-deuterated buffer for the 0 minute controls. The reaction was placed in an Agilent 1290 UPLC series autosampler and at 1, 8, 60, 180, and 1440 minutes 10 uL were withdrawn. LCMS analysis for intact protein analysis were conducted using a Phenomenex Onyx Monolithic C18 reverse phase column (100 x 2 mm) at 1°C using a flow rate of 300 μL/min under the following conditions: 1.0 min, 10% B; 1.0−4.0 min, 10−70% B; 4.0−4.5 min, 70-90% B; 4.5-6.0 min, 10% B; solvent A = 0.1% FA (Sigma) in water (Thermo-Fisher) and solvent B = 0.1% FA in ACN (Thermo-Fisher), or a Phenomenex Luna C5 column (30 x 2mm), 1°C at 200 μL/min with the following ramp: 0 min, 10% B; 1.0−6.0 min, 10−90% B; 6.0-7.0 min, 10% B. Data for both experiments were acquired in positive mode. Electrospray settings were as follows: nebulizer set to 5.0 bar, drying gas at 7.0 L/min, drying temperature at 200 °C,

and capillary voltage at 4.5 kV. The capillary exit was set at 150V, skimmer 1 at 60V, hexapole 1 at 26V, hexapole RF at 350 Vpp, and skimmer 2 at 22V. Data analysis was carried out using Bruker DataAnalysis with the MaximumEntropy plug-in. Each reaction was performed in triplicate. Statistically significant differences were determined using a two tailed students t-test with a $p<0.05$ being considered significant. All experiments were repeated a minimum of 3 times. Apo Csy was used as a baseline to normalize exchange patterns between conditions

Peptide Level HDX-MS

As described for intact HDX-MS, protein was diluted 1:10 into deuterated buffer. Ten µL were removed from the reaction mixture at 0.25, 3, 30, 180, and 1800 minutes and quenched by diluting 1:6 into 0.75% formic acid (FA, Sigma) and 0.25 mg/mL porcine pepsin (Sigma) at pH 2.5 on ice. The sample was then digested for two minutes on ice with vortexing every 30 seconds. After digestion, the samples were flash frozen and stored in liquid nitrogen until LC-MS analysis. Each reaction was performed in triplicate.

LC-MS analysis was carried out as described in Patterson *et al.(171).* Briefly, the analysis was completed on a 1290 UPLC series chromatography stack (Agilent Technologies) coupled directly to a 6538 UHD Accurate-Mass QTOF LC/MS mass spectrometer (Agilent Technologies). Before electrospray−time-of-flight (ESI-TOF) analysis, peptides were separated on a reverse phase (RP) column (Phenomenex Onyx Monolithic C18 column, 100 × 2 mm) at 1 °C using a flow rate of 500 µL/min under the following conditions: 1.0 min, 5% B; 1.0−9.0 min, 5−45% B; 9.0−11.8 min, 45-95% B;

11.80−12.0 min, 5% B; solvent A = 0.1% FA (Sigma) in water (Thermo-Fisher) and solvent B = 0.1% FA in acetonitrile (ACN, Thermo-Fisher). Data were acquired at 2 Hz s−1 over the scan range 50−1700 m/z in positive mode. Electrospray settings were as follows: nebulizer set to 3.7 bar, drying gas at 8.0 L/min, drying temperature at 350 °C, and capillary voltage at 3.5 kV. Data analysis was carried out at previously described *(106,109)* using MassHunter Qualitative Analysis (Agilent Technologies), Peptide Analysis Worksheet (PAWs, ProteoMetrics LLC), SearchGUI v3.3.16 *(108)*, PeptideShaker v 1.16.42 *(107)*, HDExaminer v 2.5.1 (Sierra Analytics), and visualized using UCSF Chimera *(154)*.

Differential Scanning Fluorimetry

Differential scanning fluorimetry was performed with a final reaction volume of 30 µL. Four experimental conditions were tested: one with Csy only protein (final concentration of 0.323 µM), one with Csy and AcrIF9 protein (final concentration of 0.457 µM), one with Csy and AcrIF2 (final concentration 0.475 µM), and another with Csy and ssDNA (final concentration of 0.523 µM). Each reaction was buffered in 100 mM Citrate Phosphate Buffer (pH 7.5) and 2.5 µL of 50x stock SPYRO orange dye (Invitrogen). Protein only and dye only controls were also tested. The samples were then loaded into the Rotor-Gene Q instrument (Qiagen) reading SYPRO orange absorbance at 570 nm as temperature was ramped up to 95°C from 25°C at 1°C per minute. Data was then plotted onto a graph of temperature vs relative fluorescence to observe the melting curve.

Results

Structural models of Acr bound Csy complexes reveal small differences in conformation compared to the unbound complex *(162,166)*. Based on previous work by our group and others, conformational change is critical to CRISPR function *(109)*. These two pieces of information led us to hypothesize that anti-CRISPRs alter the stability and dynamics of Csy in a process of allosteric inhibition. To test this, HDX-MS was used to measure the strength and stability of the hydrogen bonding network of Csy, while DSF quantified thermal stability with and without AcrIF9 or AcrIF2 bound.

Intact Protein HDX-MS Shows Global Differences in Protein Dynamics Upon Acr Binding

To probe for changes in exchange of the Csy complex on a global scale, intact protein HDX-MS was performed on AcrIF9 and AcrIF2 bound Csy and compared to the unbound form. Each complex was incubated in deuterated buffer over a time course ranging from one to 1,440 minutes. The reaction was quenched by injection into the acidic solvent stream of the LC-MS system. Deuterium uptake for each subunit of the Csy complex was calculated from the deconvoluted protein mass at each timepoint. The difference in uptake between conditions was calculated and mapped onto the structure (**Figure 5.2**). This facilitated a comparison between deuterium uptake in the Acr bound and unbound Csy complex.

The intact protein HDX-MS showed differences in exchange between the AcrIF9 and the AcrIF2 bound forms of Csy. In the AcrIF9 bound form, we observed no significant difference in deuterium exchange at the head (Cas6f, except for the 8-minute

time point) or tail (Cas5 and Cas8f except for the 8-minute time point) of the Csy complex compared to the unbound form. However, there was a significant increase in exchange in Cas7 along the backbone of the complex ($p<0.05$ from 1-60-minutes, **Figure 5.2E**). Cas7 continues to exchange throughout the HDX time course. In the AcrIF2 bound form of Csy, we observed no significant difference in deuterium uptake at the head of the complex (Cas6f). There was a significant decrease in deuterium exchange in the tail of the complex (Cas5 early time points and Cas8f at all time points **Figure 5.2C and 5.2E**) and the backbone (Cas7, all time points are statistically significant, **Figure 5.2F**). The decrease in deuterium exchange is present throughout the time course in Cas7 and is most pronounced at early time points in Cas8.

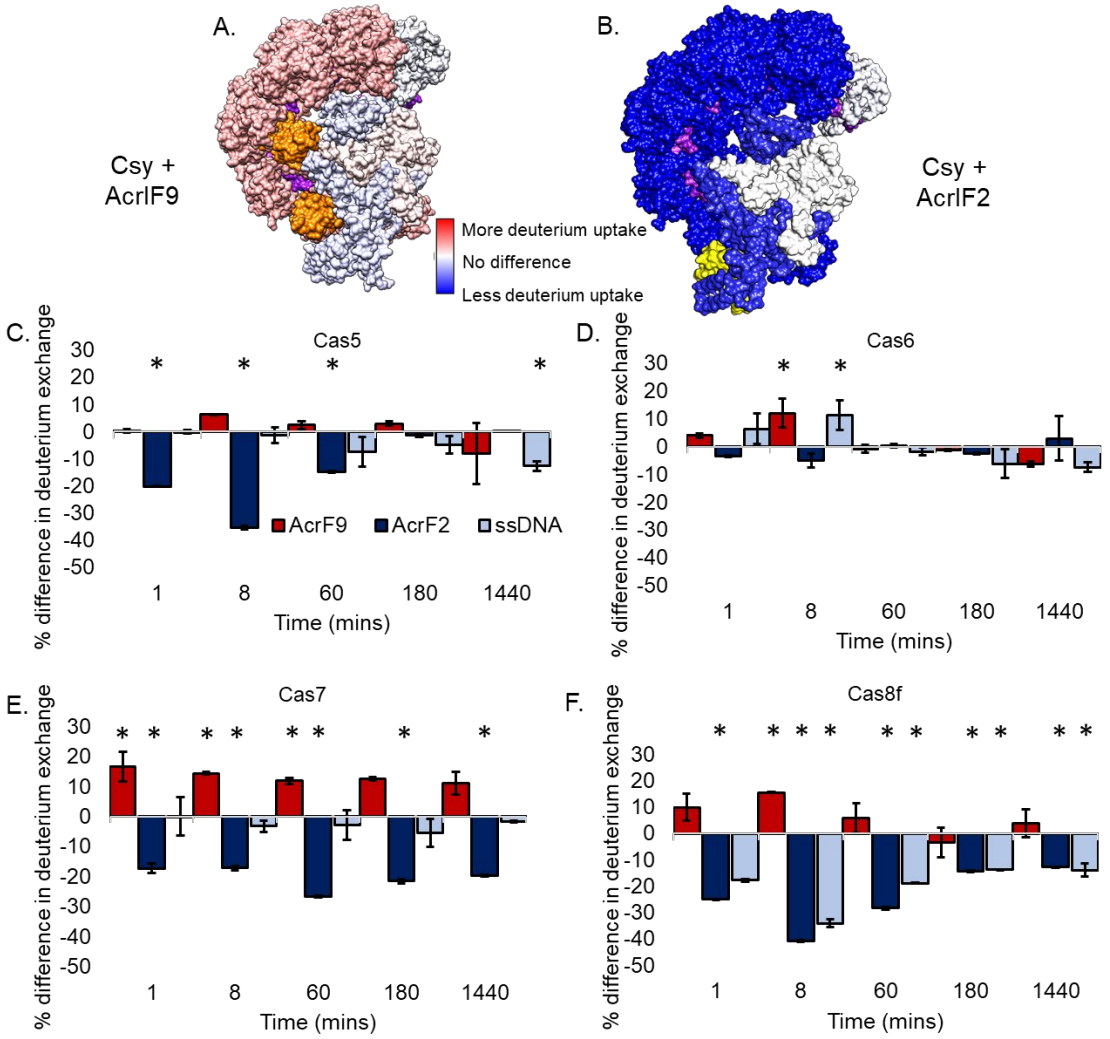

Figure 5.2. Protein level HDX shows differences in dynamics of the Csy complex when different anti-CRISPRs are bound. Intact HDX data from the 180-minute time point mapped to the AcrIF9 (A) bound structure and the AcrIF2 (B) bound structure. Areas in red show more deuterium exchange when the Acr is bound, while areas in blue show less deuterium exchange when the Acr is bound. Regions in white show no difference in exchange. Acrs are shown in orange (AcrIF9) and yellow (AcrIF2) and crRNA is shown in purple. C. The difference in deuterium uptake in either the AcrIF9 (red), AcrIF2 (dark blue), or ssDNA bound condition (light blue) compared to unbound Csy for the Cas5 subunit. Positive values indicate an increase in deuterium exchange when either the Acr or ssDNA is bound, negative values indicate a decrease in exchange. Error bars represent the standard deviation of three replicates. * indicates a significant p-value <0.05. D. Differences in deuterium exchange for the Cas6 subunit. Colors are the same as panel C.

E. Differences in deuterium exchange for the Cas7 subunit. AcrIF9 bound Csy shows a more dynamic Cas7 backbone while the AcrIF2 bound Csy shows a Cas7 backbone with a more stable secondary structure. F. Differences in deuterium exchange for the Cas8 subunit.

Peptide Level HDX-MS Shows Differences in Protein Dynamics Upon Acr Binding

To gain a higher resolution view of the changes in deuterium exchange, peptide level HDX-MS was performed on the Csy complex alone and when bound to AcrIF9 or AcrIF2. Concentrated complex was diluted into deuterated buffer and allowed to exchange for 0.5 to 1,440 minutes. The reaction was quenched with formic acid followed by pepsin digestion and LC-MS analysis. The Csy ribonucleoprotein complex is pushing the limits of HDX-MS, with four different proteins totaling >1,300 amino acid residues. Unequivocal assignment of peptides is challenging in such cases. Only peptides that could be uniquely assigned based on accurate mass and MS/MS were reported. We tracked exchange in >100 peptides across five time points, n=3. The peptides provided coverage from the head to the tail of the complex, with Cas7 having >60% sequence coverage. Deuterium uptake per peptide was calculated and used to generate heat maps showing exchange across the protein sequence.

The peptide-level differences in deuterium uptake between Csy alone and when bound by AcrIF9 or AcrIF2 were mapped on to the structure (**Figure 5.3**). The AcrIF9 bound form of Csy showed regions of increased deuterium uptake (near the crRNA) and decreased deuterium uptake (regions distal from the crRNA) (**Figure 5.3**). The increase in exchange near the crRNA is intriguing as it is present from the head to the tail of the complex with the highest increases occurring in the Cas7 backbone. While some of the

increases in exchange are modest, many of these regions show an increase in deuterium uptake of over 10% at multiple time points. In the peptides analyzed in this experiment, only three peptides within Cas7 showed bimodal behavior (99-113, 235-248, and 285-312). All other peptides showed unimodal isotopic distributions and differences in these peptides were mapped to the structure. This shows that in peptides near the crRNA the observed increase in exchange is present throughout the Csy backbone and is not localized just to the subunits that interact with the Acr.

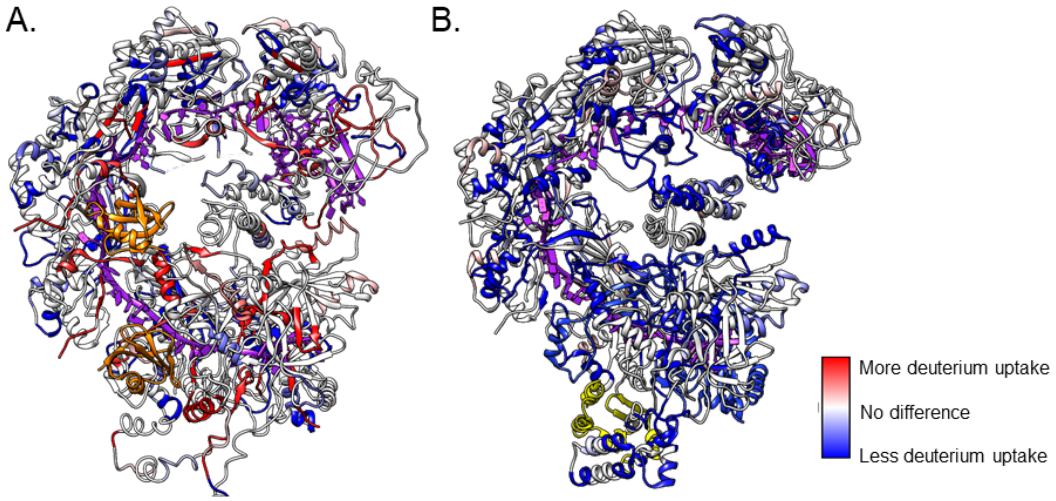

Figure 5.3. Peptide level HDX shows differences in dynamics of the Csy complex when different anti-CRISPRs are bound. Peptide level HDX data from the 180-minute time point mapped to the AcrIF9 (A) bound structure or the AcrIF2 (B) bound structure. Areas in red show more deuterium exchange when the Acr is bound, while areas in blue show less deuterium exchange when the Acr is bound. Regions in white either have no coverage or show no difference in exchange. Acrs are shown in orange (AcrIF9) or yellow (AcrIF2) and crRNA is shown in purple. AcrIF9 causes a complex change in dynamics of the Csy complex when bound while AcrIF2 shows a net stabilization of the secondary structure of the Csy complex when bound.

In accord with the intact HDX-MS data, the AcrIF2 bound Csy complex displayed a dramatically different peptide exchange profile (**Figure 5.3**). While there

were peptides with increased exchange upon AcrIF2 binding, the prevailing outcome was a significant decrease in exchange across the complex at all time points. The decrease in exchange was greater than 10% for most peptides and in many cases occurs distal from the AcrIF2 binding site.

ssDNA Binding Shows a Decrease in HDX Throughout the Csy Complex

Although AcrIF2 is a DNA mimic, the structure of Csy with AcrIF2 bound is different from dsDNA bound Csy. This prompted us to test if there were differences in stability or dynamics as well. For this experiment, a single stranded DNA (ssDNA) bound version of Csy was used. Our reasoning was that target ssDNA bound complex would undergo any associated allosteric changes, without exchange being confounded by non-target strand (displaced strand) DNA binding on the surface of Csy *(157)*.

Intact protein and peptide level HDX-MS was performed on the Csy complex bound to ssDNA. In the intact HDX-MS, we saw a significant decrease in exchange for Cas8f at the tail of the complex ($p<0.03$ at 8 minutes, **Figure 5.4B and Figure 5.2F**). The difference in exchange between the ssDNA bound Csy complex and the unbound Csy complex is statistically significant beginning with the 8-minute time point. In this instance, the Cas8f subunit takes up less deuterium upon binding of ssDNA. In the Cas7 backbone of the complex we also observed a modest decrease in exchange (**Figure 5.2E** and **Figure 5.4B**). With the peptide level HDX we observed a few areas of increased deuterium exchange upon ssDNA binding, but a majority of the peptides analyzed showed a decrease in exchange when compared to unbound Csy (**Figure 5.4A**). The net protection from exchange in the Cas7 and Cas8f subunits is redolent of the protection

from exchange observed in the AcrIF2 bound form of Csy, however, AcrIF2 protection is more significant.

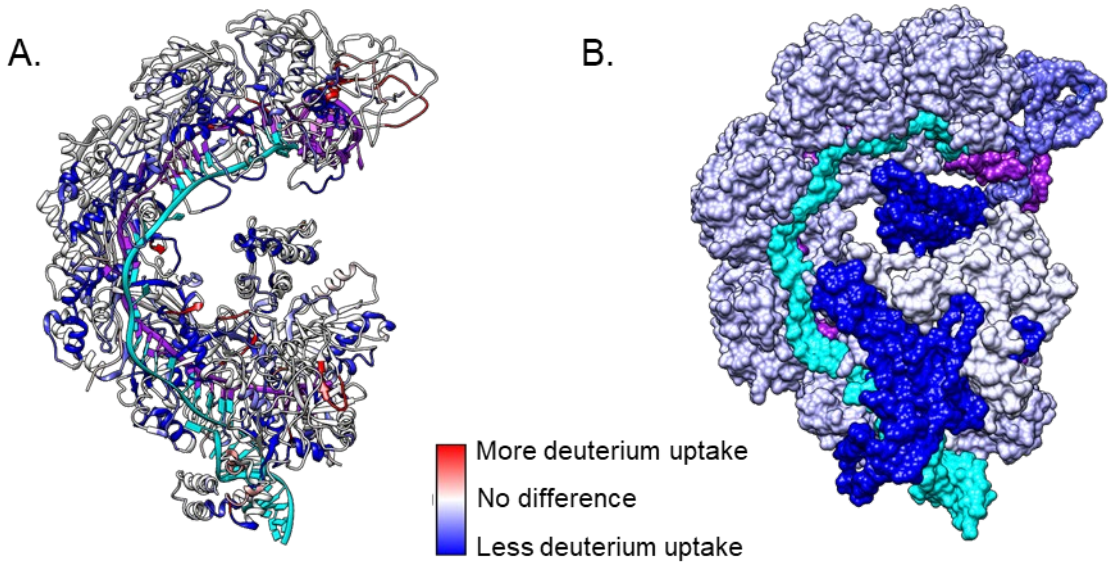

Figure 5.4. Peptide level and protein level HDX shows decreased HDX upon the binding of ssDNA. Peptide HDX data from the 180-minute time point mapped to the ssDNA bound structure (A). Protein level HDX from the 180-minute time point mapped to the ssDNA bound structure (B). Areas in red show more deuterium exchange when ssDNA is bound, while areas in blue show less deuterium exchange when ssDNA is bound. Regions in white show no difference in exchange. crRNA is shown in purple, and DNA is shown in cyan (dsDNA bound structure is used as the template for the data, PDBID: 6B44). The majority of analyzed peptides showed a decrease in exchange, especially near the crRNA backbone. Cas8f showed an overall decrease in exchange, and the Cas7 backbone showed a small decrease in exchange. This leads to the conclusion that the binding of ssDNA leads to a net stabilization of the hydrogen bonding networks, similar to what occurs when AcrIF2 is bound.

Differences in Csy Thermal Stability Upon Acr Binding

The HDX data shows changes in deuterium exchange in the AcrIF9 and AcrIF2 bound forms of Csy. This suggests that there are differences in the conformational stability and/or dynamics. The most straight-forward explanation for this is that the two

anti-CRISPRs have different modes of binding. This led us to use an orthogonal test of complex stability. The thermal stability of AcrIF9 bound, AcrIF2 bound, ssDNA bound, and Csy unbound complex was measured using Differential Scanning Fluorimetry (DSF) (**Figure 5.5**). The resulting fluorescence and temperature data were used to construct a melting curve to compare thermal protein stability between conditions.

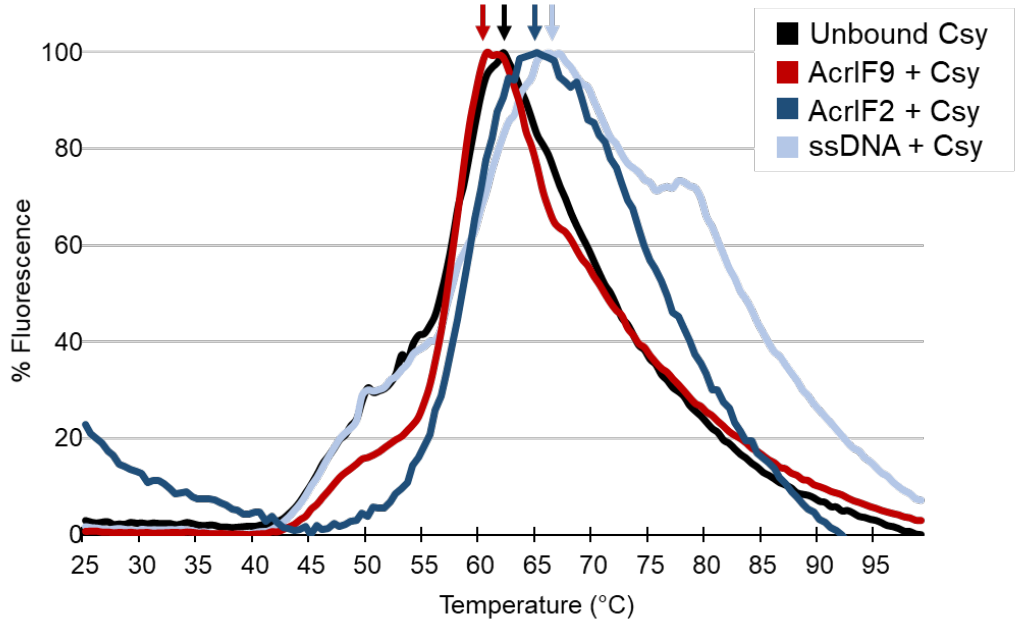

Figure 5.5. Differential Scanning Fluorometry data for the unbound (black), ssDNA (light blue), AcrIF9 (red), and AcrIF2 bound (dark blue) Csy complex. Arrows show the maximum of each peak. The maximum was defined as the melting temperature for each complex. The ssDNA and AcrIF2 bound conditions show increases in melting temperature compared to the unbound Csy complex, whereas the AcrIF9 bound condition shows a slight decrease in melting temperature when compared to the unbound Csy condition. The melting temperature differences between AcrIF2 and AcrIF9 indicate the Csy complex stability increases and decreases, respectively, upon Acr binding.

The maximum of each curve was used to determine the melting temperature (Tm) (**Figure 5.5, arrows**). With unbound Csy as a control, the ssDNA bound condition has a notably higher Tm (4ºC higher), showing that ssDNA binding thermally stabilizes the

Csy complex (**Figure 5.5, arrows**). AcrIF9 bound Csy, on the other hand, has a similar Tm as unbound Csy indicating that the global thermal stability of the complex is unaltered upon AcrIF9 binding. However, the AcrIF2 bound form shows an increase in Tm compared to unbound Csy (3ºC difference), signaling that the Csy complex becomes thermally stabilized upon AcrIF2 binding. The DSF results for AcrIF2 mirror the observations from the HDX-MS experiments.

The presence of shoulders on a melting curve rules out a simple two state system and is indicative of local unfolding events *(172)*. In the AcrIF9 bound complex, the shoulder at 50ºC is less pronounced than in the unbound and the ssDNA bound conditions. This region is further stabilized in the AcrIF2 bound condition as the shoulder is not present. Based on the intact HDX data, this shoulder could represent Cas5 or Cas7. Both subunits show similar HDX between the ssDNA and unbound Csy conditions, an increase in exchange in the AcrIF9 bound form as well as a decrease in exchange in the AcrIF2 bound form (**Figure 5.2**). The lack of a shoulder and the symmetric shape of the curve suggest that the AcrIF2 bound Csy complex unfolds cooperatively.

There is a shoulder in the ssDNA bound Csy complex at ~ 80ºC, which is well above the Tm. The ssDNA bound form of Csy is the only one that has a distinct shoulder above the Tm, which implies that a component remains folded even after most of the complex melts. The melting temperature of the ssDNA-crRNA interaction is around 65ºC. This suggests that there are strong interactions between the ssDNA and Csy.

When taken together, the DSF and HDX results offer considerable insight to Csy:ligand interactions. The DSF data reports on the relative strength of the non-

covalent interactions in the complex. HDX provides information on the strength of the hydrogen bonding network present in the secondary, tertiary, and quaternary elements of Csy. This data has allowed to assess the thermodynamic driving forces and underlying mechanism of Acr function.

Discussion

Protein-protein interactions are ultimately driven by free energy. This is formalized in the equation for Gibbs free energy (G) as, DG = DH-TDS where H is enthalpy, T temperature, and S is entropy. A reaction is favorable when the change in free energy (DG) is negative. With respect to protein:protein interactions, enthalpy will be favorable (negative) when non-covalent interactions (hydrogen bonding, van der Waals interactions, and ionic bonding) are formed between the two proteins *(173)*. As the strength of these interactions increases, the conformational space and dynamics of the protein complex decreases, lowering the entropy, or degrees of freedom within the system. Based on the equation for free energy, negative entropy makes a reaction less favorable. Thus, protein:protein interactions are a balance of enthalpic and entropic contributions. At a fundamental level, HDX provides a read out of the stability of hydrogen bonding networks, allowing it to be used as a proxy for stability and dynamics when investigating a protein under different conditions. When used judiciously, in can provide qualitative insight into the entropic and enthalpic contribution to binding. In a similar manner, melting temperatures determined by DSF can be used to assess the relative enthalpic contribution to protein binding.

AcrIF9 Binding is Entropically Stabilized

In an HDX time-course, exchange at early time points (seconds) report mostly on solvent accessibility and high frequency low amplitude motion *(20)*. Exchange at time points on the minutes to hours scale, implies lower frequency lapses in the H-bond network consistent with slow conformational change and/or protein dynamics *(21,168)*. In the case of Cas7, we observed an increase in deuterium exchange in the AcrIF9 bound form across all time points (**Figure 5.2**). This suggests that Cas7 is more solvent accessible and less stable in the AcrIF9 bound complex. A decrease in stability would mean an increase entropy for the complex which would make AcrIF9 binding favorable.

However, in a protein complex, entropy and enthalpy are in a delicate balance. As the entropy increases, the hydrogen bonding network present within the protein secondary, tertiary, or quaternary structure must weaken, leading to a less favorable enthalpy of binding. Therefore, if there is an increase in entropy in Csy upon binding of AcrIF9, one would expect either no change or a small decrease in overall complex enthalpy. Because DSF reports only on the stability of the protein complex and not the protein plus solvent, as does calorimetry, the Tm is proportional to the van't Hoff enthalpy *(174)*, or enthalpy of the complex *(172,175)*. In our DSF data, AcrIF9 bound Csy had a similar Tm and melting profile to the unbound Csy complex (**Figure 5.5**), indicating that the enthalpic contribution to stability is similar. Taken together, the HDX and DSF data support a model in which binding of AcrIF9 to the Csy complex is entropically stabilized.

AcrIF2 Binding is Enthalpically Stabilized

In the HDX data, AcrIF2 binding to the Csy complex shows an inverse trend to that of AcrIF9 binding. There is protection from deuterium exchange upon AcrIF2 binding in the Cas7 backbone as well as the tail of the complex (Cas5 and Cas8f, **Figure 5.2 and 5.3**). This decrease in deuterium exchange could be the result of a global stabilization of the hydrogen-bonding network in the AcrIF2 bound form of Csy. This implies a more favorable enthalpy, meaning stronger non-covalent interactions and a negative enthalpy upon binding. A global increase in the stability of the hydrogen bonding network would limit the conformational freedom of the complex and cause a concomitant decrease in entropy.

If AcrIF2 binding does lead to a more favorable overall enthalpy, one would predict an increase in the thermal stability of the complex. The Tm of AcrIF2 bound Csy is higher than that of Csy by about 3°C (**Figure 5.5**). This could be interpreted as stronger non-covalent interactions within the complex and a more negative enthalpic term upon binding. A more negative enthalpic term upon AcrIF2 binding is in direct agreement with the HDX where there was less exchange (longer-lived hydrogen bonding networks) in the AcrIF2 bound complex. Taken together, this data favors a model in which binding of AcrIF2 is an enthalpically stabilized reaction.

Binding of Acrs Alters the Conformational Landscape Implying a Role for Allostery

In our HDX data, we observed changes in deuterium exchange throughout the Csy complex when either Acr was bound. These global changes demonstrate that the hydrogen bonding network is altered throughout the complex, which implies that the

conformational landscape was modified upon Acr binding. The conformational landscape of a protein is a representation of the ensemble of conformations present at any given time, with the lowest energy conformations being the most populated *(7,10,176)*. The landscape contains the ligand bound conformations, however, without ligand present, they may be at a higher energy and will be infrequently populated *(10)*. Other important aspects of the landscape are the depth of a well which depicts the likelihood that the conformation will be present, and the width which depicts the relative number of conformations of similar energy. By monitoring changes in the hydrogen bond network of the Csy complex bound to AcrIF9 or AcrIF2, we were essentially tracking differences in the conformational ensemble and therefore the energy landscape (**Figure 5.6**). In the AcrIF9 bound Csy complex, we propose increased conformational freedom implying that there are a series of bound conformations with similar energy. This can be depicted as an energy landscape with a broad energy well (**Figure 5.6, red line**). In the AcrIF2 bound Csy complex, the HDX and DSF results suggest a stabilized hydrogen bond network, thus the favored ensemble would collapse into a limited set of conformations, represented by a narrow energy well (**Figure 5.6, dark blue line**). The global changes in protein dynamics may have further implications than simply stabilizing the Acr:Csy interactions.

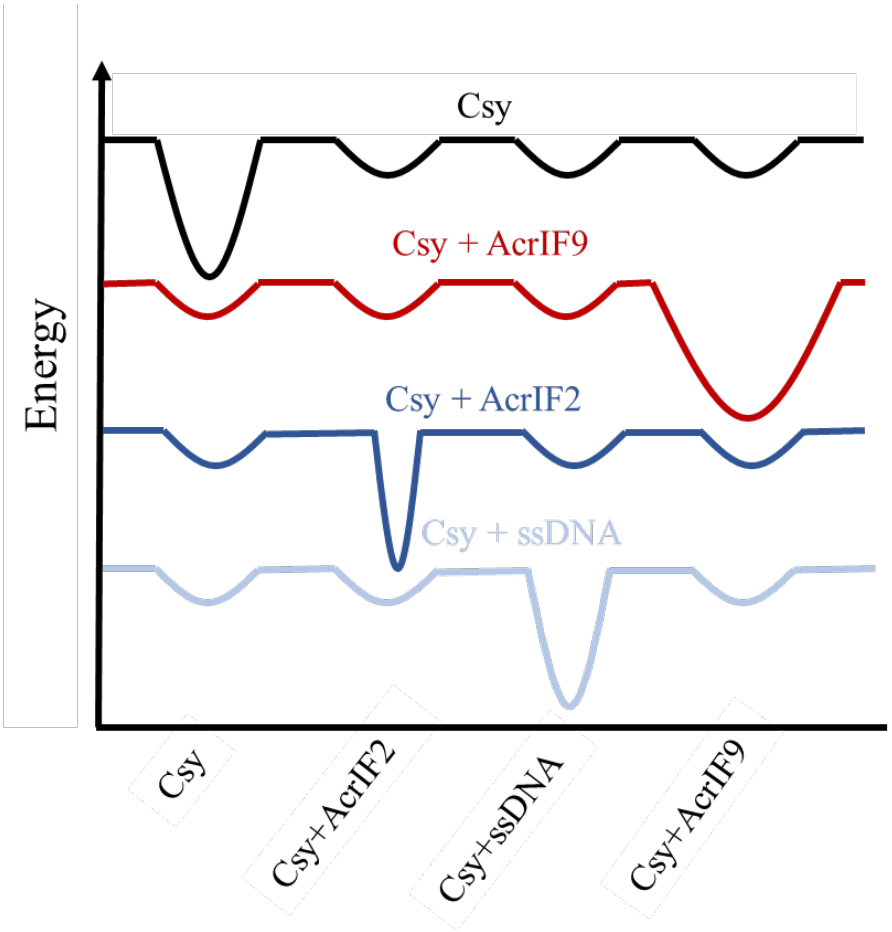

Figure 5.6. The conformational energy landscape is remodeled upon ligand binding. The hypothetical landscape shown here has four distinct populations. In the unbound state (black line) the lowest energy conformation is the unbound Csy conformation. Ligand bound conformations are present but are higher in energy. Upon AcrIF9 binding (red line), the energy landscape is remodeled such that the AcrIF9 bound conformation is the lowest energy conformation. This conformation has more favorable entropy implying that the depth of the energy well is caused by the entropy of the conformation. This well is wider than other energy wells because in increase in protein dynamics allows for the sampling of multiple conformations with the same energy. When AcrIF2 is bound (dark blue line), the conformational landscape is remodeled such that the AcrIF2 bound conformation is the most favorable. This well is stabilized by the increase in favorability of the enthalpy upon AcrIF2 binding thus causing this well to be narrower because fewer conformations can be adopted. Fewer conformations can be adopted because of the strength of the non-covalent interactions which limit protein dynamics. ssDNA binding (light blue line) shows a similar trend to AcrIF2.

The observation that there are changes in HDX distal to the Acr binding sites, suggests there is an allosteric network within Csy *(177)*. The binding of either Acr alters the hydrogen bonding network as shown here, while altering the global structure of Csy. While the structural changes are not large, there are distinct functional outcomes, as Csy no longer will bind dsDNA specifically. We posit that the functional outcomes may have more to do with the altering the thermodynamic profile of Csy, than trapping a novel conformation. This alteration in the absence of overt conformational change, can be indicative of an allosteric network *(7)*. Communication within this network begins at the binding site of the Acr to the complex and spreads from the tail to the head of the complex. This network may play a role in each ligand interaction with the Csy complex. Allosteric changes upon dsDNA binding are also present in the Type IE interference complex Cascade *(109)*. This suggests that allosteric networks play a significant role in the function of Type I CRISPR interference complexes.

AcrIF9 Interactions with DNA

Allosteric effects have an impact on the ability of Csy to bind nucleic acids. In the AcrIF2 bound Csy complex, DNA interactions are blocked, however, AcrIF9 bound Csy can still interact with dsDNA through non-specific interactions or ssDNA through sequence specific binding *(169,178)*. Non-specific binding of dsDNA to the AcrIF9 bound Csy complex occurs mainly through interactions with the two Acrs, as well as the N-terminal hook of Cas8f *(169)*. Because AcrIF9 does not bind dsDNA on its own, there must be a conformational change upon binding to Csy that facilitates affinity for dsDNA. The specificity of ssDNA binding implies that it binds directly to the crRNA template

(178). Because of this, it may be subject to Cas3 degradation while degradation of the non-specifically bound dsDNA is prevented.

In both cases of DNA binding, changes in the Cas8f subunit upon AcrIF9 binding play a role. In non-specific dsDNA binding, interactions between dsDNA and Cas8f may be facilitated through an increase in solvent accessibility or in protein dynamics. HDX shows that the Cas8f N-terminal hook has areas of increased exchange upon AcrIF9 binding (peptides covering residues 28-31, 56-65, 135-150, **Figure 5.7)** throughout the time course. These changes in exchange are in concordance with the changes observed in this region in the solved structures (**Figure 5.7B**). In the intact HDX data for Cas8f, we observed an increase in exchange in this subunit when AcrIF9 is bound at the early to middle time points (1 minute – 60 minutes, **Figure 5.2**). This increase in exchange is most pronounced at the 8-minute time point suggesting enhanced dynamics or conformational change. This change may inhibit the ability of Cas8f to perform strand separation and therefore prevent the AcrIF9 bound complex from binding dsDNA with sequence specificity. Taken together, the changes that occur in Cas8f upon AcrIF9 binding prevent the Csy complex from functioning. Because AcrIF9 does not bind in the N-terminal hook region, this illustrates the functional role of the allosteric network which is observed in the Csy complex.

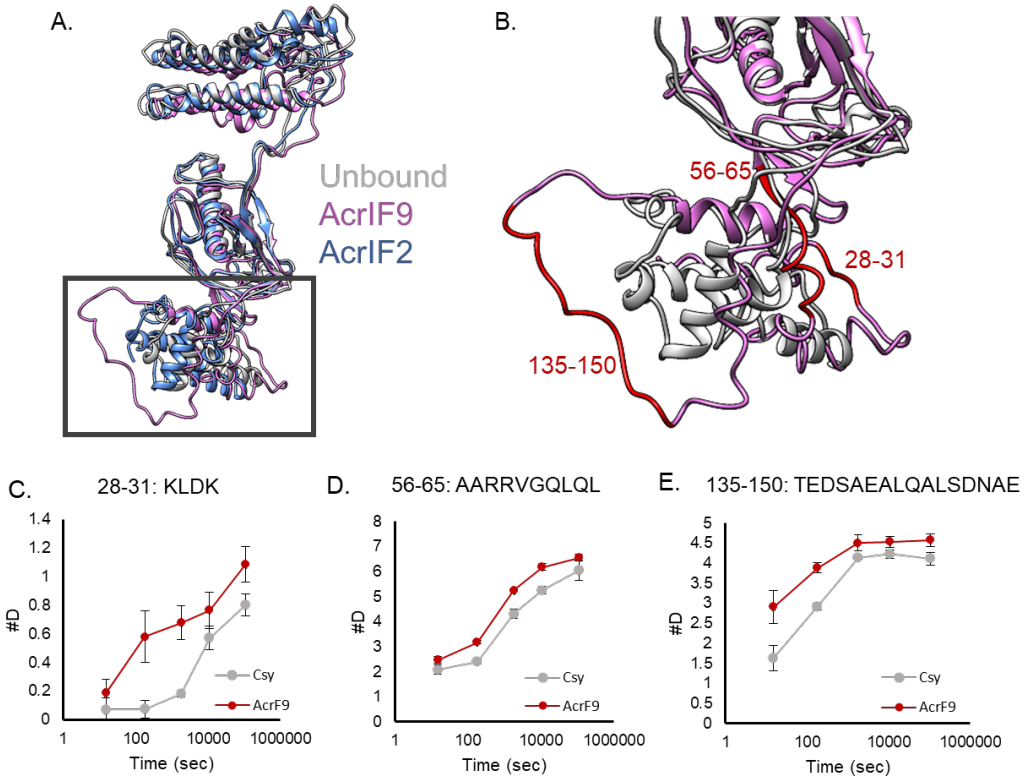

Figure 5.7. AcrIF9 changes the conformation and deuterium exchange in the Cas8f N-terminal hook region. A. Alignment of Cas8f unbound (PDBID: 6B45, white), AcrIF2 bound (PDBID: 6B47, blue), and AcrIF9 bound (PDBID: 6VQV, pink). B. A closer view of the N-terminal hook (amino acid residues 1-166) comparing the AcrIF9 bound structure (pink) to the unbound structure (grey). In this region three peptides had increased deuterium exchange in the AcrIF9 bound complex (red). C-E. Uptake curves showing the deuterium uptake for peptides that show differences in exchange in the N-terminal hook. Each data point represents the average of three replicates with error bars representing the standard deviation of those three replicates.

Conclusions

The biophysical analysis of anti-CRISPR binding to the Csy complex revealed distinct perturbations to the conformational landscape that were inhibitor specific. The two Acrs tested provide a striking example of the mechanistic diversity and thermodynamic options that drive biomolecular interactions. Upon AcrIF2 binding, we

observed a stabilization of the hydrogen bonding network of both the Cas8f subunit and the Cas7 backbone. Comparisons with the ssDNA bound complex, shows that there is a similar stabilization of Cas8f in both conditions; however, the stabilization of Cas7 in the AcrIF2 bound Csy complex is more pronounced than the ssDNA form (**Figure 5.2**). This suggests that AcrIF2 binding stabilizes a conformation that prevents productive dsDNA binding. In the AcrIF9 bound form of the complex, we observed changes in deuterium uptake which indicate that the hydrogen bonding network is altered. This implies that the conformational landscape is remodeled such that the lowest energy conformation has a more dynamic Cas7 backbone (**Figure 5.2**). This could mitigate specific dsDNA binding to the AcrIF9 bound Csy complex by adding an entropic penalty that must be overcome to adopt the dsDNA bound conformation. We posit this range of thermodynamic differences in the driving force for protein binding and the underlying thermodynamic effects on mechanisms of protein complexes are not an extraordinary example of evolution, but rather a theme that is common in the regulation of protein complexes across all domains of life.

References

(155) Barrangou, R.; Marraffini, L. A. CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Mol. Cell* **2014**, *54* (2), 234–244. https://doi.org/10.1016/j.molcel.2014.03.011.

(156) Carte, J.; Wang, R.; Li, H.; Terns, R. M.; Terns, M. P. Cas6 Is an Endoribonuclease That Generates Guide RNAs for Invader Defense in Prokaryotes. *Genes Dev.* **2008**, *22* (24), 3489–3496. https://doi.org/10.1101/gad.1742908.

(157) Guo, T. W.; Bartesaghi, A.; Yang, H.; Falconieri, V.; Rao, P.; Merk, A.; Eng, E. T.; Raczkowski, A. M.; Fox, T.; Earl, L. A.; Patel, D. J.; Subramaniam, S. Cryo-

EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* **2017**, *171* (2), 414-426.e12. https://doi.org/10.1016/j.cell.2017.09.006.

(158) Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas3 Is a Single-Stranded DNA Nuclease and ATP-Dependent Helicase in the CRISPR/Cas Immune System. *EMBO J.* **2011**, *30* (7), 1335–1342. https://doi.org/10.1038/emboj.2011.41.

(159) Pawluk, A.; Davidson, A. R.; Maxwell, K. L. Anti-CRISPR: Discovery, Mechanism and Function. *Nat. Rev. Microbiol.* **2018**, *16* (1), 12–17. https://doi.org/10.1038/nrmicro.2017.120.

(160) Bondy-Denomy, J.; Garcia, B.; Strum, S.; Du, M.; Rollins, M. F.; Hidalgo-Reyes, Y.; Wiedenheft, B.; Maxwell, K. L.; Davidson, A. R. Multiple Mechanisms for CRISPR-Cas Inhibition by Anti-CRISPR Proteins. *Nature* **2015**, *526* (7571), 136–139. https://doi.org/10.1038/nature15254.

(161) Liu, L.; Yin, M.; Wang, M.; Wang, Y. Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. *Mol. Cell* **2019**, *73* (3), 611-620.e3. https://doi.org/10.1016/j.molcel.2018.11.011.

(162) Chowdhury, S.; Carter, J.; Rollins, M. C. F.; Golden, S. M.; Jackson, R. N.; Hoffmann, C.; Nosaka, L.; Bondy-Denomy, J.; Maxwell, K. L.; Davidson, A. R.; Fischer, E. R.; Lander, G. C.; Wiedenheft, B. Structure Reveals Mechanisms of Viral Suppressors That Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **2017**, *169* (1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012.

(163) Dong, L.; Guan, X.; Li, N.; Zhang, F.; Zhu, Y.; Ren, K.; Yu, L.; Zhou, F.; Han, Z.; Gao, N.; Huang, Z. An Anti-CRISPR Protein Disables Type V Cas12a by Acetylation. *Nat. Struct. Mol. Biol.* **2019**, *26* (4), 308–314. https://doi.org/10.1038/s41594-019-0206-1.

(164) Borges, A. L.; Davidson, A. R.; Bondy-Denomy, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu. Rev. Virol.* **2017**, *4* (1), 37–59. https://doi.org/10.1146/annurev-virology-101416-041616.

(165) Pawluk, A.; Staals, R. H. J.; Taylor, C.; Watson, B. N. J.; Saha, S.; Fineran, P. C.; Maxwell, K. L.; Davidson, A. R. Inactivation of CRISPR-Cas Systems by Anti-CRISPR Proteins in Diverse Bacterial Species. *Nat. Microbiol.* **2016**, *1* (8), 1–6. https://doi.org/10.1038/nmicrobiol.2016.85.

(166) Zhang, K.; Wang, S.; Li, S.; Zhu, Y.; Pintilie, G. D.; Mou, T.-C.; Schmid, M. F.; Huang, Z.; Chiu, W. Inhibition Mechanisms of AcrF9, AcrF8, and AcrF6 against Type I-F CRISPR–Cas Complex Revealed by Cryo-EM. *Proc. Natl. Acad. Sci.* **2020**, *117* (13), 7176–7182. https://doi.org/10.1073/pnas.1922638117.

(167) Konermann, L.; Pan, J.; Liu, Y. H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/c0cs00113a.

(168) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/C0CS00113A.

(169) Hirschi, M.; Lu, W.-T.; Santiago-Frangos, A.; Wilkinson, R.; Golden, S. M.; Davidson, A. R.; Lander, G. C.; Wiedenheft, B. AcrIF9 Tethers Non-Sequence Specific DsDNA to the CRISPR RNA-Guided Surveillance Complex. *Nat. Commun.* **2020**, *11* (1), 2730. https://doi.org/10.1038/s41467-020-16512-1.

(170) Chowdhury, S.; Carter, J.; Rollins, M. C. F.; Golden, S. M.; Jackson, R. N.; Hoffmann, C.; Nosaka, L.; Bondy-Denomy, J.; Maxwell, K. L.; Davidson, A. R.; Fischer, E. R.; Lander, G. C.; Wiedenheft, B. Structure Reveals Mechanisms of Viral Suppressors That Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **2017**, *169* (1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012.

(171) Patterson, A.; Zhao, Z.; Waymire, E.; Zlotnick, A.; Bothner, B. Dynamics of Hepatitis B Virus Capsid Protein Dimer Regulate Assembly through an Allosteric Network. *ACS Chem. Biol.* **2020**, acschembio.0c00481. https://doi.org/10.1021/acschembio.0c00481.

(172) Saboury, A. A.; Moosavie-Movahedi, A. A. Clarification of Colormetric and Vant Hoff Enthalpies for Evaluation of Protein Transition States. *Biochem. Educ.* **1994**, *22* (4), 210–211. https://doi.org/10.1016/0307-4412(94)90013-2.

(173) Ladbury, J. E.; Chowdhry, B. Z. Sensing the Heat: The Application of Isothermal Titration Calorimetry to Thermodynamic Studies of Biomolecular Interactions. *Chem. Biol.* **1996**, *3* (10), 791–801. https://doi.org/10.1016/S1074-5521(96)90063-0.

(174) Lang, B. E.; Cole, K. D. Differential Scanning Calorimetry and Fluorimetry Measurements of Monoclonal Antibodies and Reference Proteins: Effect of Scanning Rate and Dye Selection. *Biotechnol. Prog.* **2017**, *33* (3), 677–686. https://doi.org/10.1002/btpr.2464.

(175) Horn, J. R.; Russell, D.; Lewis, E. A.; Murphy, K. P. Van't Hoff and Calorimetric Enthalpies from Isothermal Titration Calorimetry: Are There Significant Discrepancies? *Biochemistry* **2001**, *40* (6), 1774–1778. https://doi.org/10.1021/bi002408e.

(176) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science (80-.).* **1991**, *254* (5038), 1598–1603. https://doi.org/10.1126/science.1749933.

(177) Freire, E. The Propagation of Binding Interactions to Remote Sites in Proteins: Analysis of the Binding of the Monoclonal Antibody D1.3 to Lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 10118–10122. https://doi.org/10.1073/pnas.96.18.10118.

(178) Lu, W.-T.; Trost, C. N.; Müller-Esparza, H.; Randau, L.; Davidson, A. R. An Anti-CRISPR Protein Induces Strong Non-Specific DNA Binding Activity in a CRISPR-Cas Complex. *bioRxiv* **2020**, 2020.05.28.119941. https://doi.org/10.1101/2020.05.28.119941.

CHAPTER SIX

SUMMARY

Each of the systems studied in this thesis have implications in human health, be it in the form of understanding a disease state (Chapters 2 and 3) or potential applications as a biotechnological tool used for curing disease (Chapters 4 and 5). In order to tackle the problem of curing a disease, the underlying biochemical mechanism of the disease must be understood. Similarly, in order to use a biological system as a biotechnological tool, the system must be fully understood at the mechanistic level in order to abrogate off target effects. In this thesis, an investigation in the changes in protein dynamics within four systems were studied. This investigation provided valuable insights into the mechanism behind how these systems function in order to gain a fundamental understanding of each system.

In the four systems studied in this thesis, my research revealed that the stability and dynamics of the proteins changed upon ligand binding or mutation. Through high-resolution mapping and quantification of the solution phase properties, a deeper understanding of how the system functions was gained. In each system, changes in dynamics were observed at the site of the perturbation (ligand binding or mutation) as well as regions distal from the perturbation. The observation of changes in protein dynamics in distal regions led to the investigation of the role of the allosteric network within these complexes. The allosteric networks present were all determined to be instrumental in the function of the protein complexes studied. These allosteric networks are discussed in detail below.

Allostery in RPA

Replication protein A (RPA) is the most abundant ssDNA binding protein in cells and is involved in DNA recombination, replication, and repair *(23,179)*. Its involvement with the recruitment of enzymes that perform recombination, replication, and repair show how important this protein complex is in directing the traffic that surrounds these processes *(180)*. By investigating the changes in protein dynamics upon ssDNA binding, an understanding of what these protein recruitment signals may look like can begin to be developed. Furthermore, an understanding of how the ssDNA is handed off from RPA to the enzymes that act on the ssDNA can begin to be elucidated.

To understand the changes in protein dynamics that occur when ssDNA is bound to RPA, HDX-MS was performed on either unbound RPA or ssDNA bound RPA (**Figure 6.1**). Through this analysis it was determined that the RPA protein complex can be viewed as two halves: the less dynamic half (DBD-C,-D, and -E) and the more dynamic half (DBD-A and -B). In this context, dynamics refers to the residence time of the RPA complex domains on the ssDNA substrate. HDX-MS was used to determine if the protein:DNA interactions were stable or transient by examining the differences in exchange in the unbound RPA and ssDNA bound RPA across five time points spanning from 30-seconds to a day. The longer the domain resides on the substrate, the harder the hand off between RPA and other DNA interacting proteins. In the more dynamic half, the residence time for the RPA molecule on the DNA is less long-lived allowing for more interactions between the ssDNA and other binding partners.

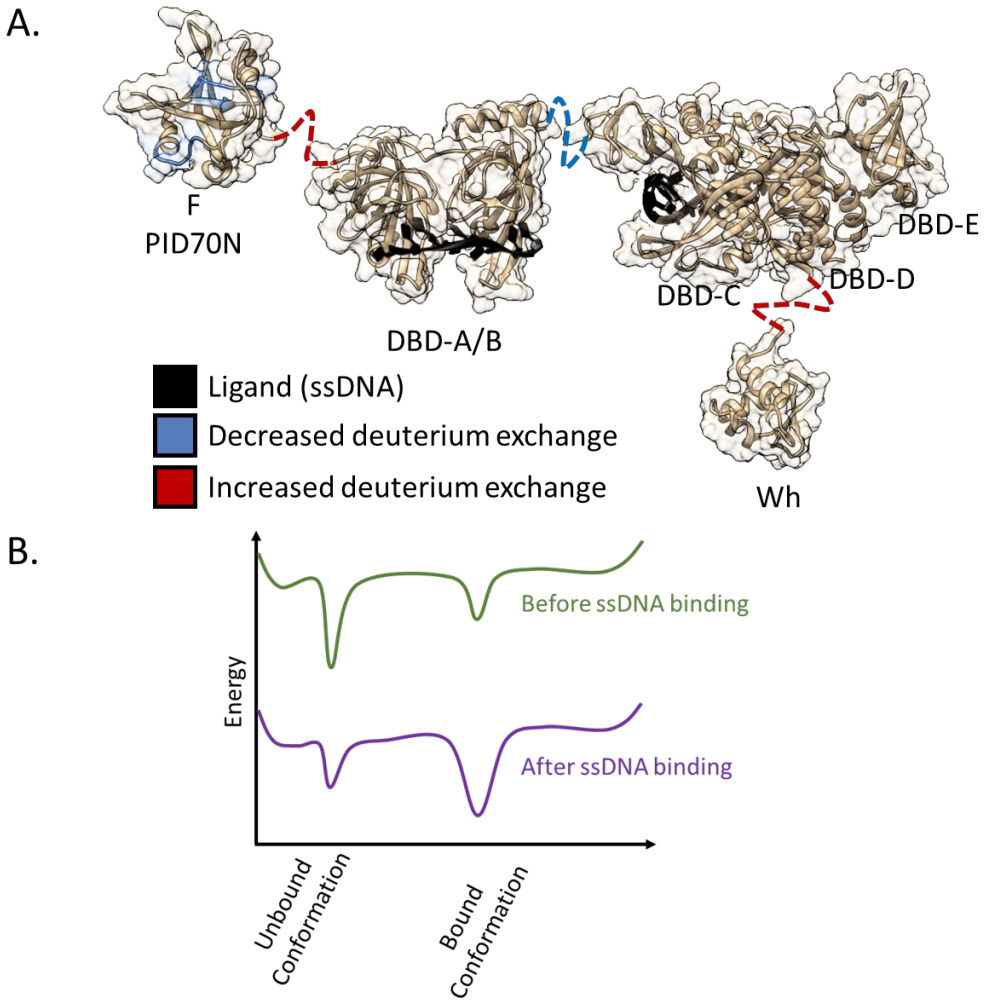

Figure 6.1 Schematic representation of the allosteric and energetic effects observed upon ssDNA binding. (A) The linker regions that connect the various DNA binding domains (DBDs) showed a difference in deuterium exchange upon ssDNA binding (red for increased exchange in the ssDNA bound form, blue for decreased exchange). This shows that the arrangement of the DBDs in relation to one another was altered upon ssDNA binding. The PID70N region showed decreased exchange upon ssDNA binding. This region is located distal to the ssDNA binding sites and is involved in the binding of other proteins to the RPA complex. This distal change in dynamics in this region shows that there is an allosteric network that connects ssDNA binding to this protein interaction site. (B) A cartoon representation of the remodeling of the energy landscape upon ssDNA binding. In the unbound state, the unbound conformation lies at the lowest energy and is the most populated conformation. As ligand binds, the equilibrium shift because the energy landscape is remodeled such that the ligand bound conformation is the most energetically favored conformation.

In the RPA system, we see changes in deuterium exchange upon ssDNA binding which are not solely localized to ssDNA interaction sites. Altered patterns of exchange are even located in domains that do not interact with DNA at all. This shows that there is a communication network located within the RPA protein complex. Upon ssDNA binding, increased deuterium exchange in the PID^{70N} domain (a non-DNA interacting domain) as well as increased exchange in the linker regions both at the 30-second time point imply an increase in solvent accessibility in these regions. This increase in solvent accessibility implies a decrease in secondary structural elements. For the disordered linker regions, this decrease propensity to adopt a stable secondary structure is not surprising because the architecture of the RPA complex is remodeled onto the ssDNA substrate which leads to differing associations between domains. By changing how the domains are organized, the linkers that connect these domains would likely be affected. However, changes in the PID^{70N} region would not necessarily have to be affected by this rearrangement of the DBDs.

The PID^{70N} region is an important protein interaction domain. Since this is where RPA interacts with other protein binding partners, it would make sense for this domain to be linked to ssDNA interacting domains in order to receive a signal that ssDNA has been productively bound. This signal may not alter the structure of the PID^{70N}, but it does alter the dynamics of this region. This functional change in dynamics in a region distal from the ssDNA binding site is a characteristic of allostery, thus, showing that allostery plays a role in RPA complex function.

Allostery in HBV

Another system that has a human health impact, is Hepatitis B Virus (HBV). Even though there is a vaccine for HBV, it is still of huge health concern causing over 800,000 deaths a year globally *(72)*. Currently, there is no cure for a chronic HBV infection which can cause cirrhosis, liver failure or hepatocellular carcinoma *(72)*. In order to discover a cure for HBV, an understanding of the viral lifecycle needs to be gained in order to identify druggable steps to prevent the replication of the virus. One of these viral lifecycle stages is capsid assembly. If we can understand how the capsid assembles, we can potentially disrupt this process and be one step closer to curing HBV.

To gain a deeper understanding of how the HBV capsid assembles, we investigated the role of protein dynamics in the assembly process. To do this, we compared the protein dynamics, through the use of HDX-MS, of four capsid protein (Cp149) dimers to determine the important sub-domains of the protein regarding capsid assembly (**Figure 6.2**). In the Y132A mutants of these dimers (hCp149 and wcCp149), assembly is blocked by the mutation. Comparing the protein dynamics of the mutant and the wildtype dimers led to the determination that the spike region plays an important role in capsid assembly.

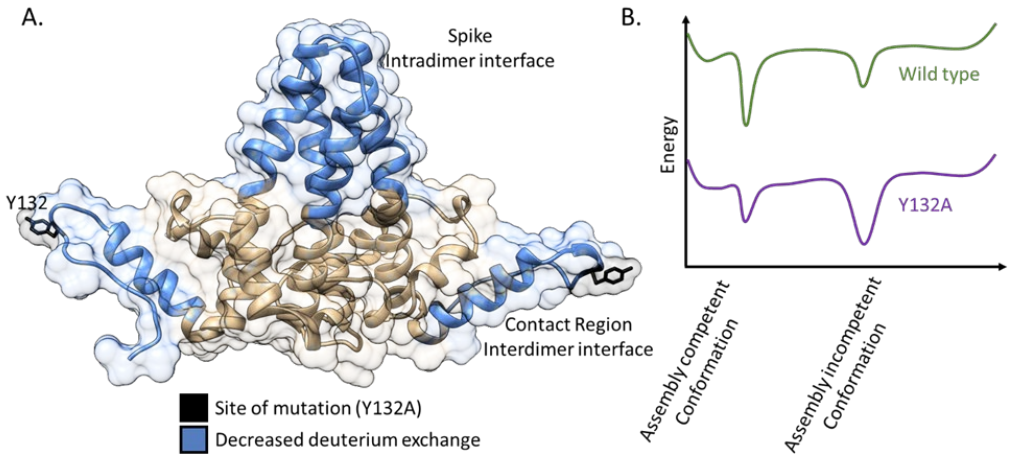

Figure 6.2. Representation of the effects of the Y132A mutation on the protein dynamics of the Cp149 dimer as well as the energy landscape. (A) Schematic representation of the subdomains within a representative HBV dimer that are affected by the mutation of Y132A. The main changes in deuterium uptake occur in the contact region and the spike region. In both regions there is an increase in deuterium exchange upon mutation of Y132 to an A. This increase in exchange is indicative of increased protein dynamics within these regions. The spike is of particular interest because this region is located on the opposite side of the monomer from the mutation location. (B) In the wild type dimer, the energy landscape shows that the assembly competent conformation is more prevalent in the population distribution than the assembly incompetent conformation. Upon mutation of Y132 to an A, this landscape is remodeled such that the assembly incompetent conformation is the conformation that a majority of the dimers are in. The low population of assembly competent dimers in the Y132A mutant, make it statistically unlikely for a capsid nucleation event to occur, thus blocking assembly.

Even though the spike is located as far from the contact region (inter-dimer interface) as possible within the protein structure, it was determined that changes in protein dynamics in this region are correlated to capsid assembly. In the capsid assembly incompetent mutants, this region was more dynamic than in the wild type Cp149. In each of the mutants (hY132A and wY132A) the intra-dimer interface was less stable than the interface in the wildtype dimer. This led to the conclusion that the stability of the secondary structural elements, in this case the alpha-helices that make up the spike tip (intra-dimer interface), is important to capsid assembly.

Assembly of the HBV capsid begins with the association of dimers through interactions in the contact region of both dimers (inter-dimer interface). The spike is located at the intra-dimer interface that is on the opposite side of the monomeric Cp149 from the contact region or the inter-dimer interface. The importance of this region, which is distal from the site of the mutation, in capsid assembly shows that there is an allosteric network within the Cp149. This allosteric network is not apparent in the static structures of each of the studied dimers, however, changes in protein dynamics upon mutation of the tyrosine at position 132 to an alanine highlights the changes in the longevity of the hydrogen-bonding network that makes up the helical structure of the spike.

In the wild type dimer, this tyrosine, which is a bulky amino acid residue, may sterically lessen the likelihood of adopting an assembly incompetent conformation, thus, causing a majority of the population of the Cp dimer to be in an assembly competent conformation. By mutating this residue to a small amino acid residue, alanine, the energy landscape of the dimer is altered such that the most energetically favorable conformation is an assembly incompetent conformation (or series of conformations) where the spike secondary structure is less long lived. By stabilizing this more dynamic conformation (or closely related conformations), the assembly process is not observed. Increased dynamics at the spike reduces the likelihood of assembly even though the dimer:dimer interface is distal to the region affected.

This shows that allosteric networks exist even in the absence of structural change or ligand binding. By perturbing the protein dynamics within the Cp149 protein, even though the structure of the protein does not change drastically, the function of the protein

was interrupted. The capsid protein is a structural protein whose function is to assemble into a T=4 viral capsid. By altering the protein dynamics of the spike tip, the capsid assembly process was interrupted, and the protein was rendered nonfunctional.

Allostery in Cascade

CRISPR systems are of interest currently because of their potential as gene editing tools *(181,182)*. Gene editing has huge implications in treating disease caused by genetic mutations, for example sickle cell disease *(183)*. By understanding how these CRISPR systems work, their applicability as gene editing tools becomes more significant through the limiting of off target effects (editing genes that should not be edited). Type I CRISPR systems are of interest because these systems completely degrade large portions of DNA. This has implications in disease states where whole genes or chromosomes need to be deleted to cure the disease, or as potential anti-microbials *(184)*. By understanding how these systems work at a fundamental level, we can begin to use these tools to combat diseases that have no cure.

To begin filling in the gaps of knowledge that existed around the mechanistic function of the Type IE interference system Cascade, changes in protein dynamics upon binding of the dsDNA substrate to the complex were investigated. This analysis was performed to determine key interaction sites between the displaced strand of DNA and the Cascade complex as well as key recruitment signal/s for the helicase/endonuclease Cas3. The path of the displaced strand of DNA was determined by looking for areas of decreased HDX within the complex. It was hypothesized that interactions with the DNA would limit solvent accessibility as well as stabilize secondary structural elements. The

pathway of the displaced strand was determined and tracked from the tail to the head of the complex.

It was hypothesized that a Cas3 recruitment signal would only be present when the Cascade complex was productively bound to dsDNA. This recruitment signal would probably have an increase in solvent accessibility or a change in dynamics upon binding of the target dsDNA. To investigate this, we looked for areas of increased deuterium exchange. The regions of most interest were located at the tail of the Cascade complex in the Cse1 subunit. This was the known interaction site for Cas3. Upon investigation, it was determined that H1 in the Cse1 subunit played an important role in Cas3:Cascade interactions (**Figure 6.3**).

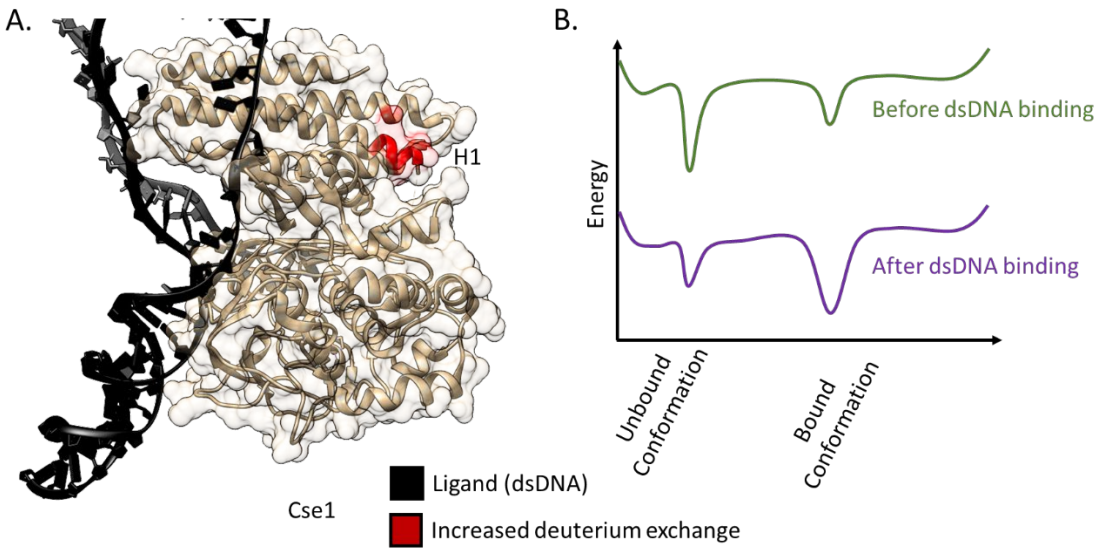

Figure 6.3 Schematic representation of the allosteric network within Cse1 of the Cascade complex and its effects on the energy landscape. (A) Upon binding of dsDNA (black ribbon) to the Cascade complex, the amount of deuterium taken up by the H1 region of Cse1 is increased. This increase in deuterium uptake implies that this region becomes more dynamic. In Chapter 4 of this thesis, it was shown that this H1 region is involved in the recruitment of the helicase/endonuclease Cas3 which degrades the foreign DNA bound to the Cascade complex. Because this H1 region is located across the protein from the dsDNA binding site, the information from dsDNA binding must be transmitted

through the protein using an allosteric network. (B) Binding of dsDNA to the Cascade complex remodels the energy landscape so that the ligand bound conformation is favored over the unbound conformation within the ensemble.

H1 is of interest as a key recruitment site for Cas3 because this helix is located across Cse1 from the dsDNA interaction site. Binding of dsDNA to the Cascade complex alters the dynamics of H1 showing the presence of an allosteric network within Cse1 allowing for communication between the dsDNA interaction site and this distal helix. This communication plays a direct role in the function of the Cascade complex. The main purpose of Cascade is to selectively bind matching dsDNA and recruit Cas3 to degrade the bound dsDNA. If the allosteric network did not exist within Cascade, Cas3 could not be recruited to degrade the foreign DNA and this adaptive immune system would no longer function.

The allosteric network within Cascade is not limited to just the Cse1 subunit. Changes in deuterium uptake upon dsDNA binding are propagated from the head (Cas6) to the tail (Cse1) of the complex. These distal changes in uptake were present across all time points indicating that there are changes in solvent accessibility as well in protein dynamics. Each subunit of this large ribonucleoprotein complex plays a role in the allosteric network present within Cascade. This shows that even large complexes can be functionally reliant on allostery.

Allostery in Csy

Also, of biotechnological interest are the use of anti-CRISPR proteins. Anti-CRISPR proteins can be used to act as an “off-switch” for CRISPR gene editing *(185)*.

These proteins prevent the CRISPR system from functioning as intended by acting as inhibitors. By temporally controlling the gene editing of CRISPR systems, it is possible to limit off target effects *(186)*. Anti-CRISPR proteins also have implications in the use of phage therapy *(187)*. In phage therapy, a bacteriophage is introduced to the site of a bacterial infection. By having the bacteriophage encode for anti-CRISPR proteins, the adaptive immune systems within the bacterial infection can be countered thus rendering the phage therapy more effective. In order to apply these two functions of anti-CRISPR proteins, the mechanism behind how these proteins inhibit their target CRISPR system must be understood.

To understand the mechanism behind how anti-CRISPRs inhibit the interference step of CRISPR immunity, the changes in protein dynamics upon binding of three different ligands to the Type IF CRISPR interference complex Csy were investigated. These ligands were either a small inhibitor protein or target ssDNA. Upon binding of any ligand, changes in deuterium exchange were tracked throughout the complex with these changes being interpreted as changes in dynamics (**Figure 6.4**). Changes in exchange were not limited to just interaction interfaces. This shows that allostery exists within the Csy complex, and this allosteric network is present regardless of the ligand bound to the complex.

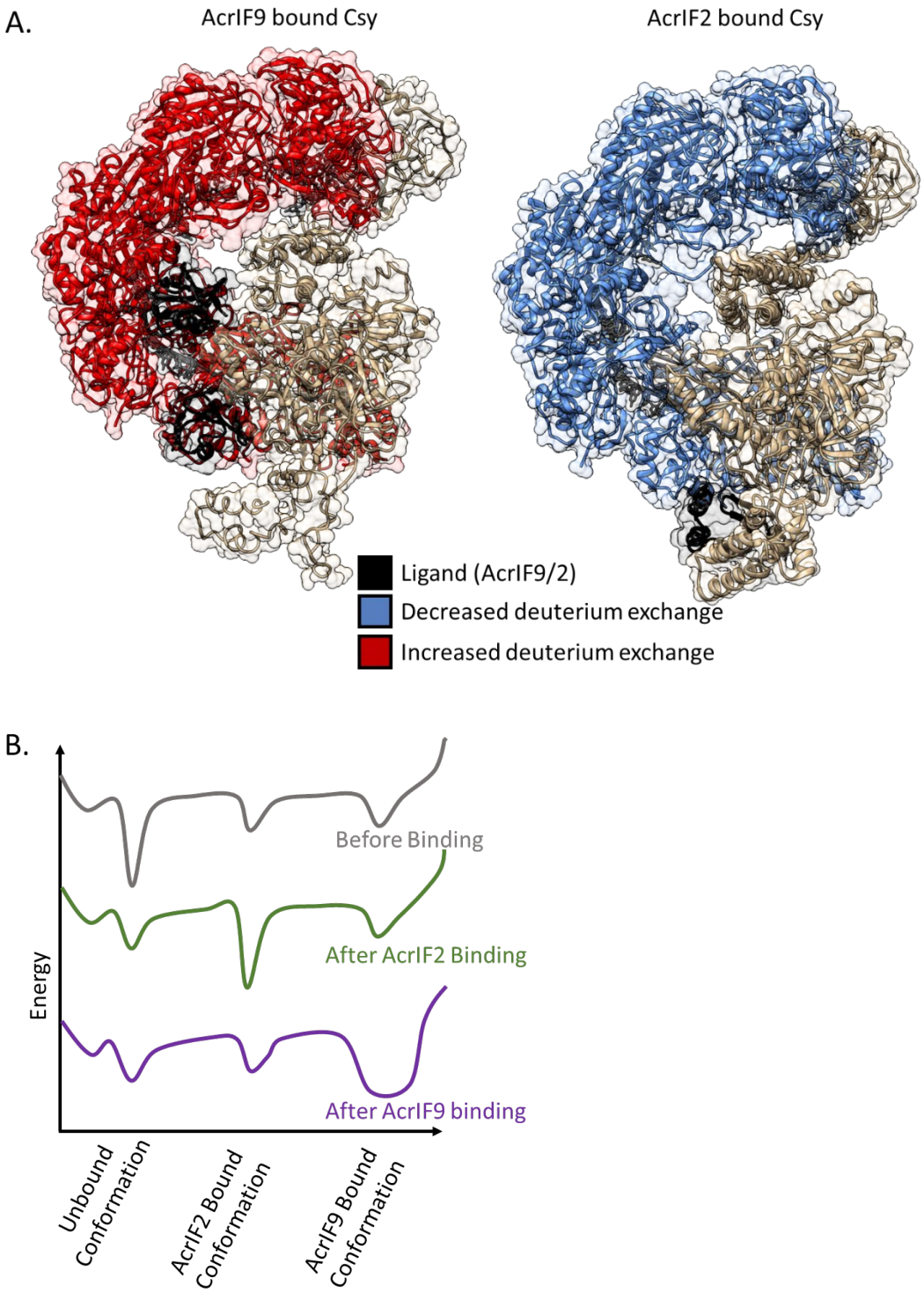

Figure 6.4 Allosteric effects and the changes in the energy landscape caused by Acr binding. (A) Schematic representation of changes in the dynamics of the Cas7 backbone of Csy upon anti-CRISPR binding. On the left, the increase of deuterium exchange (red)

when AcrIF9 (black) is bound to the Csy complex is shown. This increase in exchange was observed in the intact HDX-MS and by using peptide based HDX-MS it was determined that the increase in exchange was localized to near the crRNA (grey ribbon). On the right, the effect of the binding of AcrIF2 (black) to the Csy Cas7 dynamics is shown. There is a decrease in deuterium exchange in this region upon binding of AcrIF2 (blue). This decrease in exchange is in a different subunit of the complex from where binding occurs. This shows that there is an allosteric network present within the Csy complex, and communication through this allosteric network alters the protein dynamics upon inhibitor binding. (B) When each Acr binds, it alters the distribution of the conformational ensemble by altering the energy of the ligand bound conformation. In the AcrIF2 bound form, this interaction is enthalpically stabilized, so the energy well is a deep, narrow well. In the AcrIF9 bound form, this interaction is entropically stabilized, so the energy well is a deep but wide (multiple similar conformations) well.

Again, this network plays an important role in the function of the complex. The change in dynamics when anti-CRISPRs are bound has a functional role in preventing the Csy complex from productively binding dsDNA. AcrIF9 binding leads to an increase in protein dynamics near the crRNA thus, either providing an entropic barrier for dsDNA interactions or providing an insurmountable entropic penalty to dsDNA binding. Binding of AcrIF2, on the other hand, stabilizes the secondary structural elements of Csy such that the dsDNA bound conformation cannot be adopted. These changes in dynamics productively disrupt the function of Csy and are present across the whole of the ribonucleoprotein interference complex. By disrupting the allosteric network within Csy, either of these two Acrs are able to completely alter the function of the complex while having a small interaction interface.

Is Allostery Present in Every Protein Complex?

HDX-MS has been used to study the changes in protein dynamics in many protein systems. Changes in deuterium uptake can be observed in small proteins *(17)* as well as

large complex systems *(188)*. If allostery can be viewed as existing in the absence of structural change, then HDX-MS can be used to probe the changes in the conformational dynamics of a protein or protein complex allowing for the tracking of the allosteric network. In every protein system studied in this thesis, deuterium exchange differed at areas distal from the energetic perturbation site which translated to a change in dynamics (**Figure 6.5**). If the dynamics change then there must be a communication network that connects these distal changes from the site of perturbation. This is an allosteric network.

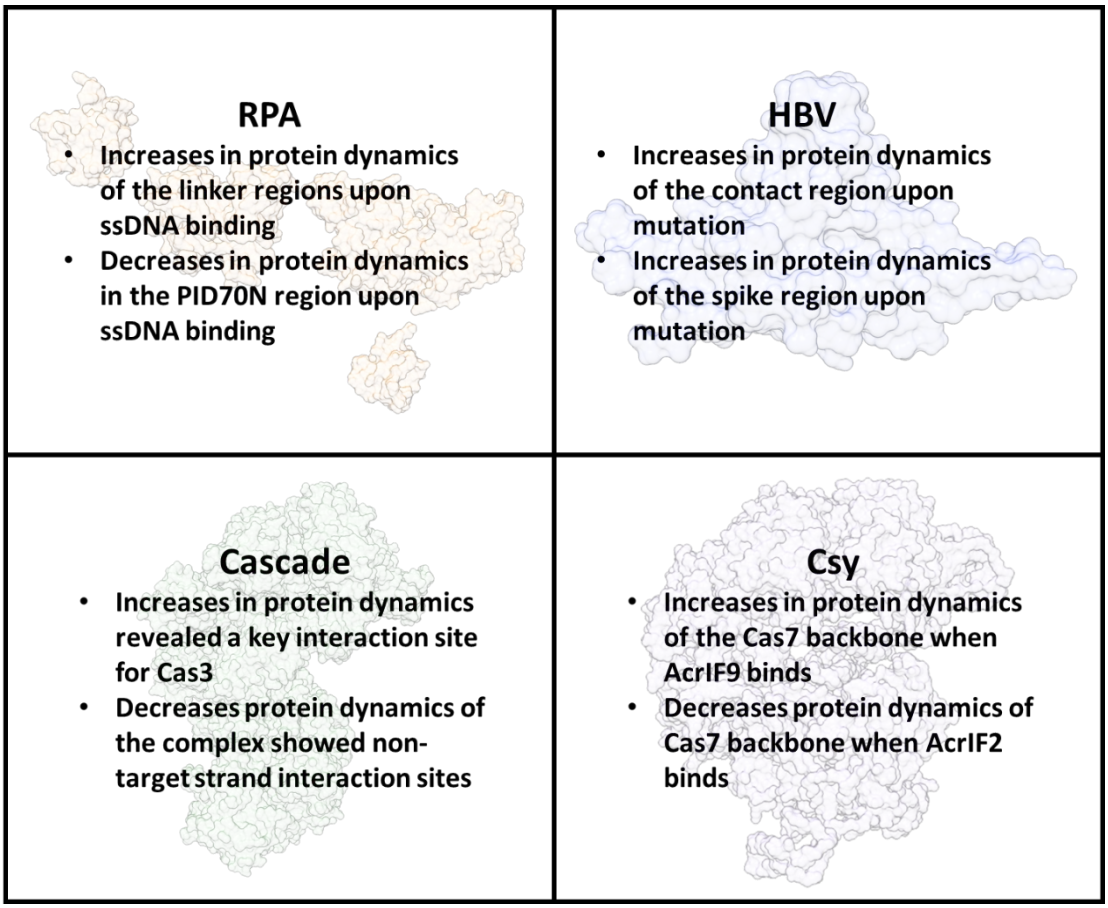

Figure 6.5 Summary of findings covered in this thesis. Four protein complexes were studied. Each complex showed the presence of an allosteric network when the changes in protein dynamics upon an energy perturbation were analyzed.

In each system I studied, I was able to define the presence of an allosteric network within the protein complex. This allosteric network played an important role in protein function, and disruption of this network altered the ability of the complex to function as intended. Allostery is present even in the studied complex systems which leads to the question, is allostery present in every protein? It would stand to reason that proteins have integral interaction networks which allow for the communication of signals across the whole structure. Protein structure is determined by covalent interactions (primary structure or disulfide bonds) as well as non-covalent interactions (secondary, tertiary, and quaternary structure). As any aspect of these structural tiers are altered, the interaction networks are adjusted to compensate for that alteration. This changes the interactions present within the protein thus changing the conformation of the protein or the dynamics of the protein (the mean conformation remains the same, but the oscillation about this mean conformation changes in either amplitude or frequency). Change in conformation can present itself as a structural conformational change or through a change in protein dynamics (stabilization or destabilization of non-covalent interactions). In respect to this definition of allostery, I would hypothesize that allosteric networks are present in every protein system, and upon an energetic perturbation, changes in non-covalent interactions will be present at functional areas distal from the perturbation site.

By definition, an energetic perturbation alters the energetic landscape of the protein. In the example of ligand binding, the ligand bound conformational state must lie at a lower energy than the unbound state in order for binding to be spontaneous. This shift in what is the most stable form of the protein illustrates that the energy landscape is

remodeled. For any energy perturbation, there will be either an increase or a decrease in energy input upon perturbation which will change the energic landscape of the protein. This change in energy landscape can present itself as a change in the average protein conformation or it can present itself as a change in protein dynamics. Because of this change in the energy landscape, the conformational differences will be propagated throughout the protein structure. This implies that allosteric networks will be able to be tracked in any protein if probed through an energetic perturbation.

References

(178) Lu, W.-T.; Trost, C. N.; Müller-Esparza, H.; Randau, L.; Davidson, A. R. An Anti-CRISPR Protein Induces Strong Non-Specific DNA Binding Activity in a CRISPR-Cas Complex. *bioRxiv* **2020**, 2020.05.28.119941. https://doi.org/10.1101/2020.05.28.119941.

(179) Wold, M. S. Replication Protein A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. *Annu. Rev. Biochem.* **1997**, *66*, 61–92. https://doi.org/10.1146/annurev.biochem.66.1.61.

(180) Oakley, G. G.; Patrick, S. M. Replication Protein A: Directing Traffic at the Intersection of Replication and Repair. *Front. Biosci.* **2010**, *15* (3), 883–900. https://doi.org/10.2741/3652.

(181) Kolli, N.; Lu, M.; Maiti, P.; Rossignol, J.; Dunbar, G. L. Application of the Gene Editing Tool, CRISPR-Cas9, for Treating Neurodegenerative Diseases. *Neurochem. Int.* **2018**, *112*, 187–196. https://doi.org/https://doi.org/10.1016/j.neuint.2017.07.007.

(182) Shen, S.; Loh, T. J.; Shen, H.; Zheng, X.; Shen, H. CRISPR as a Strong Gene Editing Tool. *BMB Rep.* **2017**, *50* (1), 20–24. https://doi.org/10.5483/bmbrep.2017.50.1.128.

(183) Hoban, M. D.; Lumaquin, D.; Kuo, C. Y.; Romero, Z.; Long, J.; Ho, M.; Young, C. S.; Mojadidi, M.; Fitz-Gibbon, S.; Cooper, A. R.; Lill, G. R.; Urbinati, F.; Campo-Fernandez, B.; Bjurstrom, C. F.; Pellegrini, M.; Hollis, R. P.; Kohn, D. B. CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ Cells. *Mol. Ther.* **2016**, *24* (9), 1561–1569.

https://doi.org/https://doi.org/10.1038/mt.2016.148.

(184) Hidalgo-Cantabrana, C.; Barrangou, R. Characterization and Applications of Type I CRISPR-Cas Systems. *Biochem. Soc. Trans.* **2020**, *48* (1), 15–23. https://doi.org/10.1042/BST20190119.

(185) Maxwell, K. L. The Anti-CRISPR Story: A Battle for Survival. *Mol. Cell* **2017**, *68* (1), 8–14. https://doi.org/https://doi.org/10.1016/j.molcel.2017.09.002.

(186) Zhang, F.; Song, G.; Tian, Y. Anti-CRISPRs: The Natural Inhibitors for CRISPR-Cas Systems. *Anim. Model. Exp. Med.* **2019**, *2* (2), 69–75. https://doi.org/https://doi.org/10.1002/ame2.12069.

(187) Azam, A. H.; Tanji, Y. Bacteriophage-Host Arm Race: An Update on the Mechanism of Phage Resistance in Bacteria and Revenge of the Phage with the Perspective for Phage Therapy. *Appl. Microbiol. Biotechnol.* **2019**, *103* (5), 2121–2131. https://doi.org/10.1007/s00253-019-09629-x.

(188) Bereszczak, J. Z.; Rose, R. J.; Duijn, E. van; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Epitope-Distal Effects Accompany the Binding of Two Distinct Antibodies to Hepatitis B Virus Capsids. *JACS* **2013**, *135* (17), 6504–6512. https://doi.org/10.1021/ja402023x.

REFERENCES CITED

(1) Levinthal, C. Are There Pathways for Protein Folding? *J. Chim. Phys.* **1968**, *65* (1), 44–45. https://doi.org/10.1051/jcp/1968650044.

(2) Englander, S. W.; Mayne, L. The Nature of Protein Folding Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (45), 15873–15880. https://doi.org/10.1073/pnas.1411798111.

(3) Ben-Nissan, G.; Sharon, M. Capturing Protein Structural Kinetics by Mass Spectrometry. *Chem. Soc. Rev.* **2011**, *40* (7), 3627–3637. https://doi.org/10.1039/c1cs15052a.

(4) Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. Folding Funnels, Binding Funnels, and Protein Function. *Protein Sci.* **1999**, *8* (6), 1181–1190. https://doi.org/10.1110/ps.8.6.1181.

(5) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science (80-.).* **1995**, *267* (5204), 1619–1620. https://doi.org/10.1126/science.7886447.

(6) Oliveberg, M.; Wolynes, P. G. The Experimental Survey of Protein-Folding Energy Landscapes. *Q. Rev. Biophys.* **2005**, *38* (3), 245–288. https://doi.org/10.1017/S0033583506004185.

(7) Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. The Ensemble Nature of Allostery. *Nature* **2014**, *508* (7496), 331–339. https://doi.org/10.1038/nature13001.

(8) Wiederschain, G. Y. *Membrane Protein Structure Determination. Methods and Protocols*; 2011; Vol. 76. https://doi.org/10.1134/s0006297911110113.

(9) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.* **1958**, *44* (2), 98–104. https://doi.org/10.1073/pnas.44.2.98.

(10) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5* (11), 789–796. https://doi.org/10.1038/nchembio.232.

(11) Monod, J.; Wyman, J.; Changeux, J. P. On the Nature of Allosteric Transitions: A Plausible Model. *J. Mol. Biol.* **1965**, *12* (1), 88–118. https://doi.org/10.1016/S0022-2836(65)80285-6.

(12) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41* (1), 585–609. https://doi.org/10.1146/annurev-biophys-050511-102319.

(13) Koshland, D. E.; Hamadani, K. Proteomics and Models for Enzyme Cooperativity. *J. Biol. Chem.* **2002**, *277* (49), 46841–46844. https://doi.org/10.1074/jbc.R200014200.

(14) Cooper, A.; Dryden, D. T. F. Allostery without Conformational Change - A Plausible Model. *Eur. Biophys. J.* **1984**, *11* (2), 103–109. https://doi.org/10.1007/BF00276625.

(15) Nussinov, R.; Tsai, C. J. Allostery without a Conformational Change? Revisiting the Paradigm. *Curr. Opin. Struct. Biol.* **2015**, *30*, 17–24. https://doi.org/10.1016/j.sbi.2014.11.005.

(16) Baldwin, R. L. Early Days of Protein Hydrogen Exchange: 1954-1972. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (7), 2021–2026. https://doi.org/10.1002/prot.23039.

(17) Hvidt, A.; Linderstrøm-Lang, K. Exchange of Hydrogen Atoms in Insulin with Deuterium Atoms in Aqueous Solutions. *BBA - Biochim. Biophys. Acta* **1954**, *14* (C), 574–575. https://doi.org/10.1016/0006-3002(54)90241-3.

(18) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/C0CS00113A.

(19) Al-Naqshabandi, M. A.; Weis, D. D. Quantifying Protection in Disordered Proteins Using Millisecond Hydrogen Exchange-Mass Spectrometry and Peptic Reference Peptides. *Biochemistry* **2017**, *56* (31), 4064–4072. https://doi.org/10.1021/acs.biochem.6b01312.

(20) McAllister, R. G.; Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **2015**, *54* (16), 2683–2692. https://doi.org/10.1021/acs.biochem.5b00215.

(21) Jaswal, S. S. Biological Insights from Hydrogen Exchange Mass Spectrometry. *Biochim. Biophys. Acta - Proteins Proteomics* **2013**, *1834* (6), 1188–1201. https://doi.org/10.1016/j.bbapap.2012.10.011.

(22) Sowole, M. A.; Konermann, L. Effects of Protein-Ligand Interactions on Hydrogen/Deuterium Exchange Kinetics: Canonical and Noncanonical Scenarios. *Anal. Chem.* **2014**, *86* (13), 6715–6722. https://doi.org/10.1021/ac501849n.

(23) Kim, C.; F. Paulus, B.; S. Wold, M. Interactions of Human Replication Protein A with Oligonucleotides. *Biochemistry* **2002**, *33* (47), 14197–14206. https://doi.org/10.1021/bi00251a031.

(24) Namiki, Y.; Zou, L. ATRIP Associates with Replication Protein A-Coated SsDNA through Multiple Interactions. *Proc Natl Acad Sci PNAS* **2006**, *103*, 580–585.

(25) Zou, L.; Elledge, S. J. Sensing DNA Damage through ATRIP Recognition of RPA-SsDNA Complexes. *Science (80-.).* **2003**, *300*, 1542–1548.

(26) Maréchal, A.; Zou, L. RPA-Coated Single-Stranded DNA as a Platform for Post-Translational Modifications in the DNA Damage Response. *Cell Res.* **2015**, *25* (1), 9–23. https://doi.org/10.1038/cr.2014.147.

(27) 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.; Chazin, W. J. A Key Interaction with RPA Orients XPA in NER Complexes. *Nucleic Acids Res.* **2020**, *48* (4), 2173–2188. https://doi.org/10.1093/nar/gkz1231.

(28) Fanning, E.; Klimovich, V.; Nager, A. R. A Dynamic Model for Replication Protein A (RPA) Function in DNA Processing Pathways. *Nucleic Acids Res.* **2006**, *34* (15), 4126–4137. https://doi.org/10.1093/nar/gkl550.

(29) Sugitani, N.; Chazin, W. J. Characteristics and Concepts of Dynamic Hub Proteins in DNA Processing Machinery from Studies of RPA. *Prog. Biophys. Mol. Biol.* **2015**, *117* (2–3), 206—211. https://doi.org/10.1016/j.pbiomolbio.2014.12.001.

(30) Bochkareva, E.; Korolev, S.; Bochkarev, A. The Role for Zinc in Replication Protein A. *J. Biol. Chem.* **2000**, *275* (35), 27332–27338. https://doi.org/10.1074/jbc.M000620200.

(31) Bochkareva, E.; Belegu, V.; Korolev, S.; Bochkarev, A. Structure of the Major Single-Stranded DNA-Binding Domain of Replication Protein A Suggests a Dynamic Mechanism for DNA Binding. *EMBO J.* **2001**, *20* (3), 612–618. https://doi.org/10.1093/emboj/20.3.612.

(32) Bochkareva, E.; Korolev, S.; Lees-Miller, S. P.; Bochkarev, A. Structure of the RPA Trimerization Core and Its Role in the Multistep DNA-Binding Mechanism of RPA. *EMBO J.* **2002**, *21* (7), 1855–1863. https://doi.org/10.1093/emboj/21.7.1855.

(33) Nakaya, R.; Takaya, J.; Onuki, T.; Moritani, M.; Nozaki, N.; Ishimi, Y. Identification of Proteins That May Directly Interact with Human RPA. *J. Biochem.* **2010**, *148* (5), 539—547. https://doi.org/10.1093/jb/mvq085.

(34) Dueva, R.; Iliakis, G. Replication Protein A: A Multifunctional Protein with Roles in DNA Replication, Repair and Beyond. *NAR Cancer* **2020**, *2* (3). https://doi.org/10.1093/narcan/zcaa022.

(35) Mer, G.; Bochkarev, A.; Gupta, R.; Bochkareva, E.; Frappier, L.; Ingles, C. J.; Edwards, A. M.; Chazin, W. J. Structural Basis for the Recognition of DNA Repair Proteins UNG2, XPA, and RAD52 by Replication Factor RPA. *Cell* **2000**, *103* (3), 449–456. https://doi.org/10.1016/s0092-8674(00)00136-7.

(36) Brosey, C. A.; Soss, S. E.; Brooks, S.; Yan, C.; Ivanov, I.; Dorai, K.; Chazin, W. J. Functional Dynamics in Replication Protein A DNA Binding and Protein Recruitment Domains. *Structure* **2015**, *23* (6), 1028–1038. https://doi.org/10.1016/j.str.2015.04.008.

(37) Brosey, C. A.; Yan, C.; Tsutakawa, S. E.; Heller, W. T.; Rambo, R. P.; Tainer, J. A.; Ivanov, I.; Chazin, W. J. A New Structural Framework for Integrating Replication Protein A into DNA Processing Machinery. *Nucleic Acids Res.* **2013**, *41* (4), 2313–2327. https://doi.org/10.1093/nar/gks1332.

(38) Iftode, C.; Borowiec, J. A. 5' --> 3' Molecular Polarity of Human Replication Protein A (HRPA) Binding to Pseudo-Origin DNA Substrates. *Biochemistry* **2000**, *39* (39), 11970–11981. https://doi.org/10.1021/bi0005761.

(39) de Laat, W. L.; Appeldoorn, E.; Sugasawa, K.; Weterings, E.; Jaspers, N. G.; Hoeijmakers, J. H. DNA-Binding Polarity of Human Replication Protein A Positions Nucleases in Nucleotide Excision Repair. *Genes Dev.* **1998**, *12* (16), 2598–2609. https://doi.org/10.1101/gad.12.16.2598.

(40) Bae, S. H.; Bae, K. H.; Kim, J. A.; Seo, Y. S. RPA Governs Endonuclease Switching during Processing of Okazaki Fragments in Eukaryotes. *Nature* **2001**, *412* (6845), 456–461. https://doi.org/10.1038/35086609.

(41) Sugawara, N.; Haber, J. E. Characterization of Double-Strand Break-Induced Recombination: Homology Requirements and Single-Stranded DNA Formation. *Mol. Cell. Biol.* **1992**, *12* (2), 563–575. https://doi.org/10.1128/mcb.12.2.563.

(42) Bastin-Shanower, S. A.; Brill, S. J. Functional Analysis of the Four DNA Binding Domains of Replication Protein A. The Role of RPA2 in SsDNA Binding. *J. Biol. Chem.* **2001**, *276* (39), 36446–36453. https://doi.org/10.1074/jbc.M104386200.

(43) Caldwell, C. C.; Spies, M. Dynamic Elements of Replication Protein A at the Crossroads of DNA Replication, Recombination, and Repair. *Crit. Rev. Biochem. Mol. Biol.* **2020**, *55* (5), 482–507. https://doi.org/10.1080/10409238.2020.1813070.

(44) Nguyen, B.; Sokoloski, J.; Galletto, R.; Elson, E. L.; Wold, M. S.; Lohman, T. M. Diffusion of Human Replication Protein A along Single-Stranded DNA. *J. Mol. Biol.* **2014**, *426* (19), 3246–3261. https://doi.org/10.1016/j.jmb.2014.07.014.

(45) Antony, E.; Lohman, T. M. Dynamics of E. Coli Single Stranded DNA Binding (SSB) Protein-DNA Complexes. *Semin. Cell Dev. Biol.* **2019**, *86*, 102–111. https://doi.org/10.1016/j.semcdb.2018.03.017.

(46) Cai, L.; Roginskaya, M.; Qu, Y.; Yang, Z.; Xu, Y.; Zou, Y. Structural Characterization of Human RPA Sequential Binding to Single-Stranded DNA Using SsDNA as a Molecular Ruler. *Biochemistry* **2007**, *46* (28), 8226–8233. https://doi.org/10.1021/bi7004976.

(47) Walther, A. P.; Gomes, X. V; Lao, Y.; Lee, C. G.; Wold, M. S. Replication Protein A Interactions with DNA. 1. Functions of the DNA-Binding and Zinc-Finger Domains of the 70-KDa Subunit. *Biochemistry* **1999**, *38* (13), 3963–3973. https://doi.org/10.1021/bi982370u.

(48) Wyka, I. M.; Dhar, K.; Binz, S. K.; Wold, M. S. Replication Protein A Interactions with DNA: Differential Binding of the Core Domains and Analysis of the DNA Interaction Surface. *Biochemistry* **2003**, *42* (44), 12909–12918. https://doi.org/10.1021/bi034930h.

(49) Arunkumar, A. I.; Stauffer, M. E.; Bochkareva, E.; Bochkarev, A.; Chazin, W. J. Independent and Coordinated Functions of Replication Protein A Tandem High Affinity Single-Stranded DNA Binding Domains. *J. Biol. Chem.* **2003**, *278* (42), 41077–41082. https://doi.org/10.1074/jbc.M305871200.

(50) Salas, T. R.; Petruseva, I.; Lavrik, O.; Saintomé, C. Evidence for Direct Contact between the RPA3 Subunit of the Human Replication Protein A and Single-Stranded DNA. *Nucleic Acids Res.* **2009**, *37* (1), 38–46. https://doi.org/10.1093/nar/gkn895.

(51) Bochkareva, E.; Kaustov, L.; Ayed, A.; Yi, G.-S.; Lu, Y.; Pineda-Lucena, A.; Liao, J. C. C.; Okorokov, A. L.; Milner, J.; Arrowsmith, C. H.; Bochkarev, A. Single-Stranded DNA Mimicry in the P53 Transactivation Domain Interaction with Replication Protein A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15412—15417. https://doi.org/10.1073/pnas.0504614102.

(52) Pokhrel, N.; Caldwell, C. C.; Corless, E. I.; Tillison, E. A.; Tibbs, J.; Jocic, N.; Tabei, S. M. A.; Wold, M. S.; Spies, M.; Antony, E. Dynamics and Selective Remodeling of the DNA-Binding Domains of RPA. *Nat. Struct. Mol. Biol.* **2019**, *26* (2), 129–136. https://doi.org/10.1038/s41594-018-0181-y.

(53) Yates, L. A.; Aramayo, R. J.; Pokhrel, N.; Caldwell, C. C.; Kaplan, J. A.; Perera, R. L.; Spies, M.; Antony, E.; Zhang, X. A Structural and Dynamic Model for the Assembly of Replication Protein A on Single-Stranded DNA. *Nat. Commun.* **2018**, *9* (1), 5447. https://doi.org/10.1038/s41467-018-07883-7.

(54) van Erp, P. B. G.; Patterson, A.; KANT, R.; Berry, L.; Golden, S. M.; Forsman, B. L.; Carter, J.; Jackson, R. N.; Bothner, B.; Wiedenheft, B. Conformational Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guide Cascade Complex. *ACS Chem. Biol.* **2017**, acschembio.7b00649. https://doi.org/10.1021/acschembio.7b00649.

(55) Berry, L.; Poudel, S.; Tokmina-Lukaszewska, M.; Colman, D. R.; Nguyen, D. M. N.; Schut, G. J.; Adams, M. W. W.; Peters, J. W.; Boyd, E. S.; Bothner, B. H/D Exchange Mass Spectrometry and Statistical Coupling Analysis Reveal a Role for Allostery in a Ferredoxin-Dependent Bifurcating Transhydrogenase Catalytic Cycle. *Biochim. Biophys. Acta - Gen. Subj.* **2018**, *1862* (1), 9–17. https://doi.org/10.1016/j.bbagen.2017.10.002.

(56) Knox, R.; Lento, C.; Wilson, D. J. Mapping Conformational Dynamics to Individual Steps in the TEM-1 β-Lactamase Catalytic Mechanism. *J. Mol. Biol.* **2018**, *430* (18 Pt B), 3311–3322. https://doi.org/10.1016/j.jmb.2018.06.045.

(57) Zheng, J.; Chang, M. R.; Stites, R. E.; Wang, Y.; Bruning, J. B.; Pascal, B. D.; Novick, S. J.; Garcia-Ordonez, R. D.; Stayrook, K. R.; Chalmers, M. J.; Dodge, J. A.; Griffin, P. R. HDX Reveals the Conformational Dynamics of DNA Sequence Specific VDR Co-Activator Interactions. *Nat. Commun.* **2017**, *8* (1), 923. https://doi.org/10.1038/s41467-017-00978-7.

(58) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* **2019**, *16* (7), 595–602. https://doi.org/10.1038/s41592-019-0459-y.

(59) Graham, B. W.; Tao, Y.; Dodge, K. L.; Thaxton, C. T.; Olaso, D.; Young, N. L.; Marshall, A. G.; Trakselis, M. A. DNA Interactions Probed by Hydrogen-Deuterium Exchange (HDX) Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Confirm External Binding Sites on the Minichromosomal Maintenance (MCM) Helicase. *J. Biol. Chem.* **2016**, *291* (24), 12467–12480. https://doi.org/10.1074/jbc.M116.719591.

(60) Zhang, J.; Li, J.; Craig, T. A.; Kumar, R.; Gross, M. L. Hydrogen-Deuterium Exchange Mass Spectrometry Reveals Calcium Binding Properties and Allosteric Regulation of Downstream Regulatory Element Antagonist Modulator (DREAM). *Biochemistry* **2017**, *56* (28), 3523—3530. https://doi.org/10.1021/acs.biochem.7b00100.

(61) Bochkarev, A.; Pfuetzner, R. A.; Edwards, A. M.; Frappier, L. Structure of the Single-Stranded-DNA-Binding Domain of Replication Protein A Bound to DNA. *Nature* **1997**, *385* (6612), 176–181. https://doi.org/10.1038/385176a0.

(62) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/nar/gky427.

(63) Pretto, D. I.; Tsutakawa, S.; Brosey, C. A.; Castillo, A.; Chagot, M.-E.; Smith, J. A.; Tainer, J. A.; Chazin, W. J. Structural Dynamics and Single-Stranded DNA Binding Activity of the Three N-Terminal Domains of the Large Subunit of Replication Protein A from Small Angle X-Ray Scattering. *Biochemistry* **2010**, *49* (13), 2880—2889. https://doi.org/10.1021/bi9019934.

(64) Corrales-Guerrero, L.; He, B.; Refes, Y.; Panis, G.; Bange, G.; Viollier, P. H.; Steinchen, W.; Thanbichler, M. Molecular Architecture of the DNA-Binding Sites of the P-Loop ATPases MipZ and ParA from Caulobacter Crescentus. *Nucleic Acids Res.* **2020**, *48* (9), 4769–4779. https://doi.org/10.1093/nar/gkaa192.

(65) Fan, J.; Pavletich, N. P. Structure and Conformational Change of a Replication Protein A Heterotrimer Bound to SsDNA. *Genes Dev.* **2012**, *26* (20), 2337–2347.

(66) Zhou, C.; Pourmal, S.; Pavletich, N. P. Dna2 Nuclease-Helicase Structure, Mechanism and Regulation by Rpa. *Elife* **2015**, *4*. https://doi.org/10.7554/eLife.09832.

(67) Binz, S. K.; Lao, Y.; Lowry, D. F.; Wold, M. S. The Phosphorylation Domain of the 32-KDa Subunit of Replication Protein A (RPA) Modulates RPA-DNA Interactions. Evidence for an Intersubunit Interaction. *J. Biol. Chem.* **2003**, *278* (37), 35584–35591. https://doi.org/10.1074/jbc.M305388200.

(68) Pokhrel, N.; Origanti, S.; Davenport, E. P.; Gandhi, D.; Kaniecki, K.; Mehl, R. A.; Greene, E. C.; Dockendorff, C.; Antony, E. Monitoring Replication Protein A (RPA) Dynamics in Homologous Recombination through Site-Specific Incorporation of Non-Canonical Amino Acids. *Nucleic Acids Res.* **2017**, *45* (16), 9413–9426. https://doi.org/10.1093/nar/gkx598.

(69) Gibb, B.; Ye, L. F.; Gergoudis, S. C.; Kwon, Y.; Niu, H.; Sung, P.; Greene, E. C. Concentration-Dependent Exchange of Replication Protein A on Single-Stranded DNA Revealed by Single-Molecule Imaging. *PLoS One* **2014**, *9* (2), e87922.

(70) Kumaran, S.; Kozlov, A. G.; Lohman, T. M. Saccharomyces Cerevisiae Replication Protein A Binds to Single-Stranded DNA in Multiple Salt-Dependent Modes. *Biochemistry* **2006**, *45* (39), 11958–11973. https://doi.org/10.1021/bi060994r.

(71) Chen, R.; Subramanyam, S.; Elcock, A. H.; Spies, M.; Wold, M. S. Dynamic Binding of Replication Protein a Is Required for DNA Repair. *Nucleic Acids Res.* **2016**, *44* (12), 5758–5772. https://doi.org/10.1093/nar/gkw339.

(72) CDA Foundation. CDA Foundation's Polaris Observatory https://cdafound.org/polaris/ (accessed Apr 9, 2020).

(73) Seeger, C.; Mason, W. S. Molecular Biology of Hepatitis B Virus Infection. *Virology* **2015**, *479–480*, 672–686. https://doi.org/10.1016/j.virol.2015.02.031.

(74) Zlotnick, A.; Cheng, N.; Conway, J. F.; Booy, F. P.; Steven, A. C.; Stahl, S. J.; Wingfield, P. T. Dimorphism of Hepatitis B Virus Capsids Is Strongly Influenced by the C-Terminus of the Capsid Protein. *Biochemistry* **1996**, *35* (23), 7412–7421. https://doi.org/10.1021/bi9604800.

(75) Bringas, R. Folding and Assembly of Hepatitis B Virus Core Protein: A New Model Proposal. *J. Struct. Biol.* **1997**, *118* (3), 189–196. https://doi.org/10.1006/jsbi.1997.3846.

(76) Crowther, F. A.; Kiselev, N. A.; Bottcher, B.; Berriman, J. A.; Borisova, G. P.; Ose, S. V; Pumpens, P. Three-Dimensional Structure of Hepatitis B Virus Core Particles Determined by Electron Cryomicroscopy. *Cell* **1994**, *77*, 943–950. https://doi.org/10.1016/0092-8674(94)90142-2.

(77) Singh, S.; Zlotnick, A. Observed Hysteresis of Virus Capsid Disassembly Is

Implicit in Kinetic Models of Assembly. *J. Biol. Chem.* **2003**, *278* (20), 18249–18255. https://doi.org/10.1074/jbc.M211408200.

(78) Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W. The Crystal Structure of the Human Hepatitis B Virus Capsid. **1999**, *3*, 771–780.

(79) Stannard, L. M.; Hodgkiss, M. Morphological Irregularities in Dane Particle Cores. *J. Gen. Virol.* **1979**, *45* (2), 509–514. https://doi.org/10.1099/0022-1317-45-2-509.

(80) Gazina, E. V.; Fielding, J. E.; Lin, B.; Anderson, D. A. Core Protein Phosphorylation Modulates Pregenomic RNA Encapsidation to Different Extents in Human and Duck Hepatitis B Viruses. *J. Virol.* **2000**, *74* (10), 4721–4728. https://doi.org/10.1128/jvi.74.10.4721-4728.2000.

(81) Porterfield, J. Z.; Dhason, M. S.; Loeb, D. D.; Nassal, M.; Stray, S. J.; Zlotnick, A. Full-Length Hepatitis B Virus Core Protein Packages Viral and Heterologous RNA with Similarly High Levels of Cooperativity. *J. Virol.* **2010**, *84* (14), 7174–7184. https://doi.org/10.1128/jvi.00586-10.

(82) Freund, S. M. V.; Johnson, C. M.; Jaulent, A. M.; Ferguson, N. Moving towards High-Resolution Descriptions of the Molecular Interactions and Structural Rearrangements of the Human Hepatitis B Core Protein. *J. Mol. Biol.* **2008**, *384* (5), 1301–1313. https://doi.org/10.1016/j.jmb.2008.10.020.

(83) Alexander, C. G.; Jurgens, M. C.; Shepherd, D. A.; Freund, S. M. V.; Ashcroft, A. E.; Ferguson, N. Thermodynamic Origins of Protein Folding, Allostery, and Capsid Formation in the Human Hepatitis B Virus Core Protein. *Proc. Natl. Acad. Sci.* **2013**, *110* (30), E2782–E2791. https://doi.org/10.1073/pnas.1308846110.

(84) Seitz, S.; Urban, S.; Antoni, C.; Böttcher, B. Cryo-Electron Microscopy of Hepatitis B Virions Reveals Variability in Envelope Capsid Interactions. *EMBO J.* **2007**, *26* (18), 4160–4167. https://doi.org/10.1038/sj.emboj.7601841.

(85) Li, H. C.; Huang, E. Y.; Su, P. Y.; Wu, S. Y.; Yang, C. C.; Lin, Y. S.; Chang, W. C.; Shih, C. Nuclear Export and Import of Human Hepatitis B Virus Capsid Protein and Particles. *PLoS Pathog.* **2010**, *6* (10). https://doi.org/10.1371/journal.ppat.1001162.

(86) Kann, M.; Schmitz, A.; Rabe, B. Intracellular Transport of Hepatitis B Virus. *World J. Gastroenterol.* **2007**, *13* (1), 39–47. https://doi.org/10.3748/wjg.v13.i1.39.

(87) Chen, C.; Wang, J. C. Y.; Pierson, E. E.; Keifer, D. Z.; Delaleau, M.; Gallucci, L.; Cazenave, C.; Kann, M.; Jarrold, M. F.; Zlotnick, A. Importin β Can Bind Hepatitis B Virus Core Protein and Empty Core-Like Particles and Induce Structural Changes. *PLoS Pathog.* **2016**, *12* (8), 1–25. https://doi.org/10.1371/journal.ppat.1005802.

(88) Selzer, L.; Kant, R.; Wang, J. C. Y.; Bothner, B.; Zlotnick, A. Hepatitis B Virus Core Protein Phosphorylation Sites Affect Capsid Stability and Transient Exposure of the C-Terminal Domain. *J. Biol. Chem.* **2015**, *290* (47), 28584–28593. https://doi.org/10.1074/jbc.M115.678441.

(89) Yeh, C. T.; Liaw, Y. F.; Ou, J. H. The Arginine-Rich Domain of Hepatitis B Virus Precore and Core Proteins Contains a Signal for Nuclear Transport. *J. Virol.* **1990**, *64* (12), 6141–6147.

(90) Liao, W.; Ou, J. H. Phosphorylation and Nuclear Localization of the Hepatitis B Virus Core Protein: Significance of Serine in the Three Repeated SPRRR Motifs. *J. Virol.* **1995**, *69* (2), 1025–1029.

(91) Kann, M.; Sodeik, B.; Vlachou, A.; Gerlich, W. H.; Helenius, A. Phosphorylation-Dependent Binding of Hepatitis B Virus Core Particles to the Nuclear Pore Complex. *J. Cell Biol.* **1999**, *145* (1), 45–55. https://doi.org/10.1083/jcb.145.1.45.

(92) Gallina, A.; Bonelli, F.; Zentilin, L.; Rindi, G.; Muttini, M.; Milanesi, G. A Recombinant Hepatitis B Core Antigen Polypeptide with the Protamine-like Domain Deleted Self-Assembles into Capsid Particles but Fails to Bind Nucleic Acids. *J. Virol.* **1989**, *63* (11), 4645–4652.

(93) Ceres, P.; Zlotnick, A. Weak Protein-Protein Interactions Are Sufficient to Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry* **2002**, *41* (39), 11525–11531. https://doi.org/10.1021/bi0261645.

(94) Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C. Hepatitis Core Antigen Produced in Escherichia Coli: Subunit Composition, Conformation Analysis, and in Vitro Capsid Assembly. *Biochemistry* **1995**, *34* (15), 4919–4932. https://doi.org/10.1021/bi00015a003.

(95) Tennant, B. C.; Gerin, J. L. The Woodchuck Model of Hepatitis B Virus Infection. *ILAR J.* **2001**, *42* (2), 89–102. https://doi.org/10.1093/ilar.42.2.89.

(96) Kukreja, A. A.; Wang, J. C.-Y.; Pierson, E.; Keifer, D. Z.; Selzer, L.; Tan, Z.; Dragnea, B.; Jarrold, M. F.; Zlotnick, A. Structurally Similar Woodchuck and Human Hepadnavirus Core Proteins Have Distinctly Different Temperature Dependences of Assembly. *J. Virol.* **2014**, *88* (24), 14105–14115. https://doi.org/10.1128/jvi.01840-14.

(97) Pierson, E. E.; Keifer, D. Z.; Kukreja, A. A.; Wang, J. C. Y.; Zlotnick, A.; Jarrold, M. F. Charge Detection Mass Spectrometry Identifies Preferred Non-Icosahedral Polymorphs in the Self-Assembly of Woodchuck Hepatitis Virus Capsids. *J. Mol. Biol.* **2016**, *428* (2), 292–300. https://doi.org/10.1016/j.jmb.2015.06.019.

(98) Asor, R.; Selzer, L.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Assembly Reactions of Hepatitis B Capsid Protein into Capsid Nanoparticles Follow a Narrow Path through a Complex Reaction Landscape. *ACS Nano* **2019**. https://doi.org/10.1021/acsnano.9b00648.

(99) Asor, R.; Schlicksup, C. J.; Zhao, Z.; Zlotnick, A.; Raviv, U. Rapidly Forming Early Intermediate Structures Dictate the Pathway of Capsid Assembly. *J. Am. Chem. Soc.* **2020**, *142* (17), 7868–7882. https://doi.org/10.1021/jacs.0c01092.

(100) Zhao, Z.; Wang, J. C.-Y.; Gonzalez-Gutierrez, G.; Venkatakrishnan, B.; Asor, R.; Khaykelson, D.; Raviv, U.; Zlotnick, A. Structural Differences between the Woodchuck Hepatitis Virus Core Protein in the Dimer and Capsid States Are Consistent with Entropic and Conformational Regulation of Assembly. *J. Virol.* **2019**, *93* (14), 1–16. https://doi.org/10.1128/jvi.00141-19.

(101) Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with a Role for Allostery in Virus Assembly. *J. Virol.* **2010**, *84* (3), 1607–1615. https://doi.org/10.1128/JVI.02033-09.

(102) Bourne, C. R.; Katen, S. P.; Fulz, M. R.; Packianathan, C.; Zlotnick, A. A Mutant Hepatitis b Virus Core Protein Mimics Inhibitors of Icosahedral Capsid Self-Assembly. *Biochemistry* **2009**, *48* (8), 1736–1742. https://doi.org/10.1021/bi801814y.

(103) Zhao, Z.; Wang, J. C.-Y.; Segura, C. P.; Hadden-Perilla, J. A.; Zlotnick, A. The Integrity of the Intradimer Interface of the Hepatitis B Virus Capsid Protein Dimer Regulates Capsid Self-Assembly. *Submitted* **2020**.

(104) Zlotnick, A.; Lee, A.; Bourne, C. R.; Johnson, J. M.; Paul, L. In Vitro Screening for Molecules That Affect Virus Capsid Assembly. *Nat. Protoc.* **2007**, *2* (3), 490–498. https://doi.org/10.1016/j.bbi.2008.05.010.

(105) Luo, M. L.; Jackson, R. N.; Denny, S. R.; Tokmina-Lukaszewska, M.; Maksimchuk, K. R.; Lin, W.; Bothner, B.; Wiedenheft, B.; Beisel, C. L. The CRISPR RNA-Guided Surveillance Complex in Escherichia Coli Accommodates Extended RNA Spacers. *Nucleic Acids Res.* **2016**, *44* (15), 7385–7394. https://doi.org/10.1093/nar/gkw421.

(106) Berry, L.; Patterson, A.; Pence, N.; Peters, J.; Bothner, B. Hydrogen Deuterium Exchange Mass Spectrometry of Oxygen Sensitive Proteins. *Bio-Protocol* **2018**, *8* (6), 1–16. https://doi.org/10.21769/BioProtoc.2769.

(107) Vaudel, M.; Burkhart, J. M.; Zahedi, R. P.; Oveland, E.; Berven, F. S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker Enables Reanalysis of MS-Derived Proteomics Data Sets: To the Editor. *Nat. Biotechnol.* **2015**, *33* (1), 22–24. https://doi.org/10.1038/nbt.3109.

(108) Barsnes, H.; Vaudel, M. SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *J. Proteome Res.* **2018**, *17* (7), 2552–2555. https://doi.org/10.1021/acs.jproteome.8b00175.

(109) Van Erp, P. B. G.; Patterson, A.; Kant, R.; Berry, L.; Golden, S. M.; Forsman, B. L.; Carter, J.; Jackson, R. N.; Bothner, B.; Wiedenheft, B. Conformational

Dynamics of DNA Binding and Cas3 Recruitment by the CRISPR RNA-Guided Cascade Complex. *ACS Chem. Biol.* **2018**, *13* (2), 481–490. https://doi.org/10.1021/acschembio.7b00649.

(110) Rayaprolu, V.; Kruse, S.; Kant, R.; Movahed, N.; Brooke, D.; Bothner, B. Fluorometric Estimation of Viral Thermal Stability. *Bio-Protocol* **2014**, *4* (15). https://doi.org/10.21769/bioprotoc.1199.

(111) Niesen, F. H.; Berglund, H.; Vedadi, M. The Use of Differential Scanning Fluorimetry to Detect Ligand Interactions That Promote Protein Stability. *Nat. Protoc.* **2007**, *2* (9), 2212–2221. https://doi.org/10.1038/nprot.2007.321.

(112) Chen, F.; Gulbakan, B.; Weidmann, S.; Fagerer, S. R.; Ibanez, A. J.; Zenobi, R. Applying Mass Spectrometry to Study Non-Covalent Biomolecule Complexes. *Mass Spectrom. Rev.* **2015**, *35* (1), 48–70. https://doi.org/10.1002/mas.21462.

(113) Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. Protein Dynamics Viewed by Hydrogen Exchange. *Protein Sci.* **2012**, *21* (7), 996–1005. https://doi.org/10.1002/pro.2081.

(114) Selzer, L.; Katen, S. P.; Zlotnick, A. The Hepatitis B Virus Core Protein Intradimer Interface Modulates Capsid Assembly and Stability. *Biochemistry* **2014**, *53* (34), 5496–5504. https://doi.org/10.1021/bi500732b.

(115) Hilmer, J. K.; Zlotnick, A.; Bothner, B. Conformational Equilibria and Rates of Localized Motion within Hepatitis B Virus Capsids. *J. Mol. Biol.* **2008**, *375* (2), 581–594. https://doi.org/10.1016/j.jmb.2007.10.044.

(116) Bereszczak, J. Z.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Assessment of Differences in the Conformational Flexibility of Hepatitis B Virus Core-Antigen and e-Antigen by Hydrogen Deuterium Exchange-Mass Spectrometry. *Protein Sci.* **2014**, *23* (7), 884–896. https://doi.org/10.1002/pro.2470.

(117) Stray, S. J.; Bourne, C. R.; Punna, S.; Lewis, W. G.; Finn, M. G.; Zlotnick, A. A Heteroaryldihydropyrimidine Activates and Can Misdirect Hepatitis B Virus Capsid Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (23), 8138–8143. https://doi.org/10.1073/pnas.0409732102.

(118) Brouns, S. J. J.; Jore, M. M.; Lundgren, M.; Westra, E. R.; Slijkhuis, R. J. H.; Snijders, A. P. L.; Dickman, M. J.; Makarova, K. S.; Koonin, E. V.; van der OOst, J. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **2013**, *321* (5891), 960–965. https://doi.org/10.1126/science.1159689.

(119) Jore, M. M.; Lundgren, M.; Van Duijn, E.; Bultema, J. B.; Westra, E. R.; Waghmare, S. P.; Wiedenheft, B.; Pul, Ü.; Wurm, R.; Wagner, R.; Beijer, M. R.; Barendregt, A.; Zhou, K.; Snijders, A. P. L.; Dickman, M. J.; Doudna, J. A.; Boekema, E. J.; Heck, A. J. R.; Van Der Oost, J.; Brouns, S. J. J. Structural Basis for CRISPR RNA-Guided DNA Recognition by Cascade. *Nat. Struct. Mol. Biol.*

2011, *18* (5), 529–536. https://doi.org/10.1038/nsmb.2019.

(120) Jackson, R. N.; Golden, S. M.; Erp, P. B. G. Van; Carter, J.; Westra, E. R.; Brouns, S. J. J.; Oost, J. Van Der; Terwilliger, T. C.; Read, R. J.; Wiedenheft, B. Crystal Structure of the CRISPR RNA– Guided Surveillance Complex from Escherichia Coli. *Science (80-.).* **2014**, *345* (6203), 1473–1479. https://doi.org/10.1126/science.1256328.

(121) Zhao, H.; Sheng, G.; Wang, J.; Wang, M.; Bunkoczi, G.; Gong, W.; Wei, Z.; Wang, Y. Crystal Structure of the RNA-Guided Immune Surveillance Cascade Complex in Escherichia Coli. *Nature* **2014**, *515* (7525), 147–150. https://doi.org/10.1038/nature13733.

(122) Mulepati, S.; Bailey, S. In Vitro Reconstitution of an Escherichia Coli RNA-Guided Immune System Reveals Unidirectional, ATP-Dependent Degradation of DNA Target. *J. Biol. Chem.* **2013**, *288* (31), 22184–22192. https://doi.org/10.1074/jbc.M113.472233.

(123) Hochstrasser, M. L.; Taylor, D. W.; Bhat, P.; Guegler, C. K.; Sternberg, S. H.; Nogales, E.; Doudna, J. A. CasA Mediates Cas3-Catalyzed Target Degradation during CRISPR RNA-Guided Interference. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (18), 6618–6623. https://doi.org/10.1073/pnas.1405079111.

(124) Westra, E. R.; van Erp, P. B. G.; Künne, T.; Wong, S. P.; Staals, R. H. J.; Seegers, C. L. C.; Bollen, S.; Jore, M. M.; Semenova, E.; Severinov, K.; de Vos, W. M.; Dame, R. T.; de Vries, R.; Brouns, S. J. J.; van der Oost, J. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol. Cell* **2012**, *46* (5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018.

(125) Hayes, R. P.; Xiao, Y.; Ding, F.; Van Erp, P. B. G.; Rajashankar, K.; Bailey, S.; Wiedenheft, B.; Ke, A. Structural Basis for Promiscuous PAM Recognition in Type I-E Cascade from E. Coli. *Nature* **2016**, *530* (7591), 499–503. https://doi.org/10.1038/nature16995.

(126) Semenova, E.; Jore, M. M.; Datsenko, K. A.; Semenova, A.; Westra, E. R.; Wanner, B.; van der Oost, J.; Brouns, S. J. J.; Severinov, K. Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence. *Proc. Natl. Acad. Sci.* **2011**, *108* (25), 10098–10103. https://doi.org/10.1073/pnas.1104144108.

(127) Jung, C.; Hawkins, J. A.; Jones, S. K.; Xiao, Y.; Rybarski, J. R.; Dillard, K. E.; Hussmann, J.; Saifuddin, F. A.; Savran, C. A.; Ellington, A. D.; Ke, A.; Press, W. H.; Finkelstein, I. J. Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **2017**, *170* (1), 35-47.e13. https://doi.org/10.1016/j.cell.2017.05.044.

(128) van Erp, P. B. G.; Jackson, R. N.; Carter, J.; Golden, S. M.; Bailey, S.;

Wiedenheft, B. Mechanism of CRISPR-RNA Guided Recognition of DNA Targets in Escherichia Coli. *Nucleic Acids Res.* **2015**, *43* (17), 8381–8391. https://doi.org/10.1093/nar/gkv793.

(129) Mulepati, S.; Héroux, A.; Bailey, S. Structural Biology. Crystal Structure of a CRISPR RNA-Guided Surveillance Complex Bound to a SsDNA Target. *Science* **2014**, *345* (6203), 1479–1484. https://doi.org/10.1126/science.1256996.

(130) Xiao, Y.; Luo, M.; Hayes, R. P.; Kim, J.; Ng, S.; Ding, F.; Liao, M.; Ke, A. Structure Basis for Directional R-Loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **2017**, *170* (1), 48-60.e11. https://doi.org/10.1016/j.cell.2017.06.012.

(131) Fineran, P. C.; Gerritzen, M. J. H.; Suárez-Diez, M.; Künne, T.; Boekhorst, J.; van Hijum, S. A. F. T.; Staals, R. H. J.; Brouns, S. J. J. Degenerate Target Sites Mediate Rapid Primed CRISPR Adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (16), E1629–E1638. https://doi.org/10.1073/pnas.1400071111.

(132) Westra, E. R.; Semenova, E.; Datsenko, K. A.; Jackson, R. N.; Wiedenheft, B.; Severinov, K.; Brouns, S. J. J. Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genet.* **2013**, *9* (9), e1003742–e1003742. https://doi.org/10.1371/journal.pgen.1003742.

(133) Datsenko, K. A.; Pougach, K.; Tikhonov, A.; Wanner, B. L.; Severinov, K.; Semenova, E. Molecular Memory of Prior Infections Activates the CRISPR/Cas Adaptive Bacterial Immunity System. *Nat. Commun.* **2012**, *3*, 945. https://doi.org/10.1038/ncomms1937.

(134) Richter, C.; Dy, R. L.; McKenzie, R. E.; Watson, B. N. J.; Taylor, C.; Chang, J. T.; McNeil, M. B.; Staals, R. H. J.; Fineran, P. C. Priming in the Type I-F CRISPR-Cas System Triggers Strand-Independent Spacer Acquisition, Bi-Directionally from the Primed Protospacer. *Nucleic Acids Res.* **2014**, *42* (13), 8516–8526. https://doi.org/10.1093/nar/gku527.

(135) Redding, S.; Sternberg, S. H.; Marshall, M.; Gibb, B.; Bhat, P.; Guegler, C. K.; Wiedenheft, B.; Doudna, J. A.; Greene, E. C. Surveillance and Processing of Foreign DNA by the Escherichia Coli CRISPR-Cas System. *Cell* **2015**, *163* (4), 854–865. https://doi.org/10.1016/j.cell.2015.10.003.

(136) Xue, C.; Whitis, N. R.; Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell* **2016**, *64* (4), 826–834. https://doi.org/10.1016/j.molcel.2016.09.033.

(137) Xue, C.; Seetharam, A. S.; Musharova, O.; Severinov, K.; Brouns, S. J. J.; Severin, A. J.; Sashital, D. G. CRISPR Interference and Priming Varies with Individual Spacer Sequences. *Nucleic Acids Res.* **2015**, *43* (22), 10831–10847. https://doi.org/10.1093/nar/gkv1259.

(138) Swarts, D. C.; Mosterd, C.; van Passel, M. W. J.; Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **2012**, *7* (4), e35888–e35888. https://doi.org/10.1371/journal.pone.0035888.

(139) Blosser, T. R.; Loeff, L.; Westra, E. R.; Vlot, M.; Künne, T.; Sobota, M.; Dekker, C.; Brouns, S. J. J.; Joo, C. Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell* **2015**, *58* (1), 60–70. https://doi.org/10.1016/j.molcel.2015.01.028.

(140) Guttman, M.; Weis, D. D.; Engen, J. R.; Lee, K. K. Analysis of Overlapped and Noisy Hydrogen/Deuterium Exchange Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (12), 1906–1912. https://doi.org/10.1007/s13361-013-0727-5.

(141) Hourdel, V.; Volant, S.; O'Brien, D. P.; Chenal, A.; Chamot-Rooke, J.; Dillies, M.-A.; Brier, S. MEMHDX: An Interactive Tool to Expedite the Statistical Validation and Visualization of Large HDX-MS Datasets. *Bioinformatics* **2016**, *32* (22), 3413–3419. https://doi.org/10.1093/bioinformatics/btw420.

(142) Engen, J. R. Analysis of Protein Conformation and Dynamics by Hydrogen/Deuterium Exchange MS. *Anal. Chem.* **2009**, *81* (19), 7870–7875. https://doi.org/10.1021/ac901154s.Analysis.

(143) Xue, C.; Whitis, N. R.; Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell* **2016**, *64* (4), 826–834. https://doi.org/10.1016/j.molcel.2016.09.033.

(144) Jackson, R. N.; van Erp, P. B.; Sternberg, S. H.; Wiedenheft, B. Conformational Regulation of CRISPR-Associated Nucleases. *Curr. Opin. Microbiol.* **2017**, *37*, 110–119. https://doi.org/10.1016/j.mib.2017.05.010.

(145) Sashital, D. G.; Jinek, M.; Doudna, J. A. An RNA-Induced Conformational Change Required for CRISPR RNA Cleavage by the Endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **2011**, *18* (6), 680–687. https://doi.org/10.1038/nsmb.2043.

(146) Szczelkun, M. D.; Tikhomirova, M. S.; Sinkunas, T.; Gasiunas, G.; Karvelis, T.; Pschera, P.; Siksnys, V.; Seidel, R. Direct Observation of R-Loop Formation by Single RNA-Guided Cas9 and Cascade Effector Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (27), 9798–9803. https://doi.org/10.1073/pnas.1402597111.

(147) Almendros, C.; Guzmán, N. M.; Díez-Villaseñor, C.; García-Martínez, J.; Mojica, F. J. M. Target Motifs Affecting Natural Immunity by a Constitutive CRISPR-Cas System in Escherichia Coli. *PLoS One* **2012**, *7* (11), e50797–e50797. https://doi.org/10.1371/journal.pone.0050797.

(148) Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A. RNA-Guided Genetic Silencing Systems in Bacteria and Archaea. *Nature* **2012**, *482* (7385), 331–338. https://doi.org/10.1038/nature10886.

(149) Mulepati, S.; Orr, A.; Bailey, S. Crystal Structure of the Largest Subunit of a

Bacterial RNA-Guided Immune Complex and Its Role in DNA Target Binding. *J. Biol. Chem.* **2012**, *287* (27), 22445–22449. https://doi.org/10.1074/jbc.C112.379503.

(150) Sashital, D. G.; Wiedenheft, B.; Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **2012**, *46* (5), 606–615. https://doi.org/10.1016/j.molcel.2012.03.020.

(151) Tay, M.; Liu, S.; Yuan, Y. A. Crystal Structure of Thermobifida Fusca Cse1 Reveals Target DNA Binding Site. *Protein Sci.* **2015**, *24* (2), 236–245. https://doi.org/10.1002/pro.2609.

(152) Wiedenheft, B.; Lander, G. C.; Zhou, K.; Jore, M. M.; Brouns, S. J. J.; Van Der Oost, J.; Doudna, J. A.; Nogales, E. Structures of the RNA-Guided Surveillance Complex from a Bacterial Immune System. *Nature* **2011**, *477* (7365), 486–489. https://doi.org/10.1038/nature10402.

(153) Rutkauskas, M.; Sinkunas, T.; Songailiene, I.; Tikhomirova, M. S.; Siksnys, V.; Seidel, R. Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep.* **2015**, *10* (9), 1534–1543. https://doi.org/10.1016/j.celrep.2015.01.067.

(154) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(155) Barrangou, R.; Marraffini, L. A. CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Mol. Cell* **2014**, *54* (2), 234–244. https://doi.org/10.1016/j.molcel.2014.03.011.

(156) Carte, J.; Wang, R.; Li, H.; Terns, R. M.; Terns, M. P. Cas6 Is an Endoribonuclease That Generates Guide RNAs for Invader Defense in Prokaryotes. *Genes Dev.* **2008**, *22* (24), 3489–3496. https://doi.org/10.1101/gad.1742908.

(157) Guo, T. W.; Bartesaghi, A.; Yang, H.; Falconieri, V.; Rao, P.; Merk, A.; Eng, E. T.; Raczkowski, A. M.; Fox, T.; Earl, L. A.; Patel, D. J.; Subramaniam, S. Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* **2017**, *171* (2), 414-426.e12. https://doi.org/10.1016/j.cell.2017.09.006.

(158) Sinkunas, T.; Gasiunas, G.; Fremaux, C.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas3 Is a Single-Stranded DNA Nuclease and ATP-Dependent Helicase in the CRISPR/Cas Immune System. *EMBO J.* **2011**, *30* (7), 1335–1342. https://doi.org/10.1038/emboj.2011.41.

(159) Pawluk, A.; Davidson, A. R.; Maxwell, K. L. Anti-CRISPR: Discovery, Mechanism and Function. *Nat. Rev. Microbiol.* **2018**, *16* (1), 12–17.

https://doi.org/10.1038/nrmicro.2017.120.

(160) Bondy-Denomy, J.; Garcia, B.; Strum, S.; Du, M.; Rollins, M. F.; Hidalgo-Reyes, Y.; Wiedenheft, B.; Maxwell, K. L.; Davidson, A. R. Multiple Mechanisms for CRISPR-Cas Inhibition by Anti-CRISPR Proteins. *Nature* **2015**, *526* (7571), 136–139. https://doi.org/10.1038/nature15254.

(161) Liu, L.; Yin, M.; Wang, M.; Wang, Y. Phage AcrIIA2 DNA Mimicry: Structural Basis of the CRISPR and Anti-CRISPR Arms Race. *Mol. Cell* **2019**, *73* (3), 611-620.e3. https://doi.org/10.1016/j.molcel.2018.11.011.

(162) Chowdhury, S.; Carter, J.; Rollins, M. C. F.; Golden, S. M.; Jackson, R. N.; Hoffmann, C.; Nosaka, L.; Bondy-Denomy, J.; Maxwell, K. L.; Davidson, A. R.; Fischer, E. R.; Lander, G. C.; Wiedenheft, B. Structure Reveals Mechanisms of Viral Suppressors That Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **2017**, *169* (1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012.

(163) Dong, L.; Guan, X.; Li, N.; Zhang, F.; Zhu, Y.; Ren, K.; Yu, L.; Zhou, F.; Han, Z.; Gao, N.; Huang, Z. An Anti-CRISPR Protein Disables Type V Cas12a by Acetylation. *Nat. Struct. Mol. Biol.* **2019**, *26* (4), 308–314. https://doi.org/10.1038/s41594-019-0206-1.

(164) Borges, A. L.; Davidson, A. R.; Bondy-Denomy, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu. Rev. Virol.* **2017**, *4* (1), 37–59. https://doi.org/10.1146/annurev-virology-101416-041616.

(165) Pawluk, A.; Staals, R. H. J.; Taylor, C.; Watson, B. N. J.; Saha, S.; Fineran, P. C.; Maxwell, K. L.; Davidson, A. R. Inactivation of CRISPR-Cas Systems by Anti-CRISPR Proteins in Diverse Bacterial Species. *Nat. Microbiol.* **2016**, *1* (8), 1–6. https://doi.org/10.1038/nmicrobiol.2016.85.

(166) Zhang, K.; Wang, S.; Li, S.; Zhu, Y.; Pintilie, G. D.; Mou, T.-C.; Schmid, M. F.; Huang, Z.; Chiu, W. Inhibition Mechanisms of AcrF9, AcrF8, and AcrF6 against Type I-F CRISPR–Cas Complex Revealed by Cryo-EM. *Proc. Natl. Acad. Sci.* **2020**, *117* (13), 7176–7182. https://doi.org/10.1073/pnas.1922638117.

(167) Konermann, L.; Pan, J.; Liu, Y. H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/c0cs00113a.

(168) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem. Soc. Rev.* **2011**, *40* (3), 1224–1234. https://doi.org/10.1039/C0CS00113A.

(169) Hirschi, M.; Lu, W.-T.; Santiago-Frangos, A.; Wilkinson, R.; Golden, S. M.; Davidson, A. R.; Lander, G. C.; Wiedenheft, B. AcrIF9 Tethers Non-Sequence Specific DsDNA to the CRISPR RNA-Guided Surveillance Complex. *Nat. Commun.* **2020**, *11* (1), 2730. https://doi.org/10.1038/s41467-020-16512-1.

(170) Chowdhury, S.; Carter, J.; Rollins, M. C. F.; Golden, S. M.; Jackson, R. N.; Hoffmann, C.; Nosaka, L.; Bondy-Denomy, J.; Maxwell, K. L.; Davidson, A. R.; Fischer, E. R.; Lander, G. C.; Wiedenheft, B. Structure Reveals Mechanisms of Viral Suppressors That Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **2017**, *169* (1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012.

(171) Patterson, A.; Zhao, Z.; Waymire, E.; Zlotnick, A.; Bothner, B. Dynamics of Hepatitis B Virus Capsid Protein Dimer Regulate Assembly through an Allosteric Network. *ACS Chem. Biol.* **2020**, acschembio.0c00481. https://doi.org/10.1021/acschembio.0c00481.

(172) Saboury, A. A.; Moosavie-Movahedi, A. A. Clarification of Colormetric and Vant Hoff Enthalpies for Evaluation of Protein Transition States. *Biochem. Educ.* **1994**, *22* (4), 210–211. https://doi.org/10.1016/0307-4412(94)90013-2.

(173) Ladbury, J. E.; Chowdhry, B. Z. Sensing the Heat: The Application of Isothermal Titration Calorimetry to Thermodynamic Studies of Biomolecular Interactions. *Chem. Biol.* **1996**, *3* (10), 791–801. https://doi.org/10.1016/S1074-5521(96)90063-0.

(174) Lang, B. E.; Cole, K. D. Differential Scanning Calorimetry and Fluorimetry Measurements of Monoclonal Antibodies and Reference Proteins: Effect of Scanning Rate and Dye Selection. *Biotechnol. Prog.* **2017**, *33* (3), 677–686. https://doi.org/10.1002/btpr.2464.

(175) Horn, J. R.; Russell, D.; Lewis, E. A.; Murphy, K. P. Van't Hoff and Calorimetric Enthalpies from Isothermal Titration Calorimetry: Are There Significant Discrepancies? *Biochemistry* **2001**, *40* (6), 1774–1778. https://doi.org/10.1021/bi002408e.

(176) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science (80-.).* **1991**, *254* (5038), 1598–1603. https://doi.org/10.1126/science.1749933.

(177) Freire, E. The Propagation of Binding Interactions to Remote Sites in Proteins: Analysis of the Binding of the Monoclonal Antibody D1.3 to Lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (18), 10118–10122. https://doi.org/10.1073/pnas.96.18.10118.

(178) Lu, W.-T.; Trost, C. N.; Müller-Esparza, H.; Randau, L.; Davidson, A. R. An Anti-CRISPR Protein Induces Strong Non-Specific DNA Binding Activity in a CRISPR-Cas Complex. *bioRxiv* **2020**, 2020.05.28.119941. https://doi.org/10.1101/2020.05.28.119941.

(179) Wold, M. S. Replication Protein A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism. *Annu. Rev. Biochem.* **1997**, *66*, 61–92. https://doi.org/10.1146/annurev.biochem.66.1.61.

(180) Oakley, G. G.; Patrick, S. M. Replication Protein A: Directing Traffic at the

Intersection of Replication and Repair. *Front. Biosci.* **2010**, *15* (3), 883–900. https://doi.org/10.2741/3652.

(181) Kolli, N.; Lu, M.; Maiti, P.; Rossignol, J.; Dunbar, G. L. Application of the Gene Editing Tool, CRISPR-Cas9, for Treating Neurodegenerative Diseases. *Neurochem. Int.* **2018**, *112*, 187–196. https://doi.org/https://doi.org/10.1016/j.neuint.2017.07.007.

(182) Shen, S.; Loh, T. J.; Shen, H.; Zheng, X.; Shen, H. CRISPR as a Strong Gene Editing Tool. *BMB Rep.* **2017**, *50* (1), 20–24. https://doi.org/10.5483/bmbrep.2017.50.1.128.

(183) Hoban, M. D.; Lumaquin, D.; Kuo, C. Y.; Romero, Z.; Long, J.; Ho, M.; Young, C. S.; Mojadidi, M.; Fitz-Gibbon, S.; Cooper, A. R.; Lill, G. R.; Urbinati, F.; Campo-Fernandez, B.; Bjurstrom, C. F.; Pellegrini, M.; Hollis, R. P.; Kohn, D. B. CRISPR/Cas9-Mediated Correction of the Sickle Mutation in Human CD34+ Cells. *Mol. Ther.* **2016**, *24* (9), 1561–1569. https://doi.org/https://doi.org/10.1038/mt.2016.148.

(184) Hidalgo-Cantabrana, C.; Barrangou, R. Characterization and Applications of Type I CRISPR-Cas Systems. *Biochem. Soc. Trans.* **2020**, *48* (1), 15–23. https://doi.org/10.1042/BST20190119.

(185) Maxwell, K. L. The Anti-CRISPR Story: A Battle for Survival. *Mol. Cell* **2017**, *68* (1), 8–14. https://doi.org/https://doi.org/10.1016/j.molcel.2017.09.002.

(186) Zhang, F.; Song, G.; Tian, Y. Anti-CRISPRs: The Natural Inhibitors for CRISPR-Cas Systems. *Anim. Model. Exp. Med.* **2019**, *2* (2), 69–75. https://doi.org/https://doi.org/10.1002/ame2.12069.

(187) Azam, A. H.; Tanji, Y. Bacteriophage-Host Arm Race: An Update on the Mechanism of Phage Resistance in Bacteria and Revenge of the Phage with the Perspective for Phage Therapy. *Appl. Microbiol. Biotechnol.* **2019**, *103* (5), 2121–2131. https://doi.org/10.1007/s00253-019-09629-x.

(188) Bereszczak, J. Z.; Rose, R. J.; Duijn, E. van; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Epitope-Distal Effects Accompany the Binding of Two Distinct Antibodies to Hepatitis B Virus Capsids. *JACS* **2013**, *135* (17), 6504–6512. https://doi.org/10.1021/ja402023x.