Abstract

Structural Studies on Mechanisms of SAMHD1 Function and Dysfunction
Joshua E. Temple
2021

Sterile α-motif and HD domain-containing protein 1 (SAMHD1) is a pivotal enzyme that maintains cellular nucleotide homeostasis. SAMHD1 forms a tetramer upon binding nucleotides at allosteric sites, activating its triphosphohydrolase activity against substrate deoxynucleoside triphosphates (dNTPs). This dNTPase activity, in combination with functions in DNA repair, cell cycle regulation, and nucleic acid binding, is important to other facets of SAMHD1 biology as a critical mediator of antiviral defense, cancer suppression, and immune regulation. An array of post-translational modifications on SAMHD1 presents another layer of functional regulation, including phosphorylation at T592 that modulates SAMHD1's many functions. Yet, precise mechanisms for SAMHD1 cellular functions remain unclear and previous efforts examining structural consequences of T592 phosphorylation are incomplete. Some clinically-relevant mutations decouple SAMHD1's intertwined activities and allow us to parse the molecular determinants of SAMHD1 functions mentioned above. Here I present a structural analysis of cancer-associated SAMHD1 mutants that, in combination with functional data, show the importance of SAMHD1 dNTPase activity to avoid cancer proliferation. Furthermore, I present the first high resolution cryo-electron microscopy (cryo-EM) structures of full-length human SAMHD1 and a T592D phosphomimetic which reveal little difference and show extreme SAM domain flexibility. Collectively, these studies provide deeper insight into determinants of SAMHD1 function.

Structural Studies on Mechanisms of SAMHD1 Function and Dysfunction

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

By
Joshua E. Temple

Dissertation Director: Yong Xiong, PhD

December 2021

Acknowledgments

First, I'd like to recognize my mother and father, Peggy and John Temple, both of whom made many sacrifices to put me through school. Their unending support was instrumental in earning this degree. Thank you to my sister, Kimber Carey, for being my best friend and encouraging me nonstop; that extends to my brother-in-law Chris, niece Kiernan, and nephew Ceallan. To my brother Cody, this one's for you. To all aunts, uncles, and adopted families and friends along the way that took care of me, thank you. I love you all.

Importantly, I'd like to thank my committee. It was an honor to learn from Dr. Yong Xiong, who is a brilliant scientist and teacher, a kind mentor, and a friend. His guidance along with the support and advice from Drs. Karin Reinisch and Walther Mothes was invaluable. I'd also like to thank Dr. Gina MacDonald, my friend and advisor from JMU, who turned my good farm work into good lab work. These and many other great scientists have helped shape my journey as a structural biologist.

I also want to send heartfelt thanks to a few groups: the Xiong lab members in its many iterations who made it easy to come to work each day; my fantastic team of doctors who took excellent care of me; and the JMU biochem crew – Courtney, Nicole, Allie, and Mike – who are still tight knit today. Other important individuals include my best friend Billy, for his endless support and motivation. I wish we met sooner. Thanks to Aspen, my oldest friend. And thank you to the best boy, Apollo, my goofy dog.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Overview of SAMHD1 structure, activity, and modifications

Nucleotides represent not only the fundamental monomeric building blocks of nucleic acids but also an essential energy source for biological reactions across all kingdoms of life. Ribonucleoside triphosphates (rNTPs) fuel RNA synthesis and take on roles as biological batteries for kinases, whereas the sole purpose of deoxyribonucleoside triphosphates (dNTPs) is to fuel DNA synthesis[1]. Because DNA synthesis is required for replication, the cell cycle is therefore a primary driver of, and is highly correlated with, dNTP levels[2]. Deviation from normal dNTP concentrations has been shown to greatly enhance mutation rates and facilitate defective DNA replication[3]. Further, host cells are not alone in usage of cellular resources. Obligate DNA-producing viruses must also parasitize host cell dNTP pools in order to replicate. Thus, sensing and controlling the balance of this common DNA fuel source is essential for maintaining host metabolism while simultaneously limiting the replication competency of viruses.

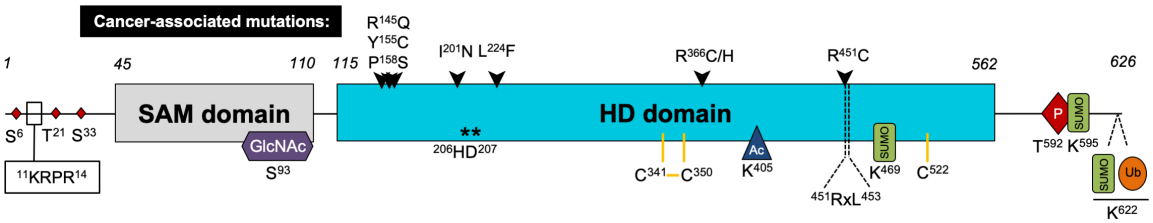

Figure 1-1. Domain schematic of human SAMHD1. Cancer mutations relevant to this dissertation are noted above. Various sites of modification or of regulatory importance as discussed in the body of the text are listed below. The N-terminal NLS is boxed. The ^{451}RxL^{453} cyclin-binding motif is noted. Small red diamonds at the N-terminus indicate phosphorylation sites of unknown consequence. GlcNAc, N-Acetylglucosamine; Ac, acetylation; SUMO, small ubiquitin-like modifier; Ub, ubiquitin; P, phosphorylation.

At all points in the cell cycle, dNTP concentrations must be tightly controlled through the concerted action of enzymes that detect, synthesize, and/or degrade dNTPs[4, 5]. Multiple cellular enzymes contribute to dNTP pool homeostasis. The enzyme ribonucleotide reductase (RNR) senses relative dNTP levels and chemically reduces the appropriate rNTPs to make dNTP precursors as needed[6]. Sterile α motif and histidine/aspartate domain containing protein 1 (SAMHD1) however is an enzyme that negatively regulates the cellular dNTP pool through its triphosphohydrolase activity, among many other independent functions[7]. Triphosphohydrolases cleave dNTP substrates into component 2'-deoxyribonucleoside (dN) and inorganic triphosphate (TP_i) products. This activity is essential and has largely been evolutionarily preserved going back to prokaryotes, with the earliest example of a SAMHD1 orthologue being found in *Escherichia coli*: the hexameric dGTP triphosphohydrolase (dGTPase) which is specific to hydrolyzing dGTP and lacks an N-terminal SAM domain[8]. SAMHD1 remains the only known eukaryotic dNTP triphosphohydrolase and instead binds all dNTPs at the catalytic site following an elegant nucleotide-coupled activation mechanism[9-11].

Domain structure of human SAMHD1 is schematized in Figure 1-1. SAMHD1 exists in equilibrium between monomeric/dimeric states and the catalytically active homotetramer, a dimer of dimers[12, 13]. Binding of either GTP[9, 14, 15] (the primary activator due to higher cellular concentrations) or dGTP[16-18] at allosteric nucleotide binding site 1 (A1) in two monomers induces dimerization. Further dNTP binding at allosteric site 2 (A2) in all subunits forms an active

tetramer capable of hydrolyzing dNTPs. Thus 8 total allosteric nucleotides act as molecular staples, each contacting three monomers, to stabilize an enzymatically-competent tetramer that binds 4 substrate dNTPs (Figure 1-2). Each substrate dNTP is coordinated with metal ions and a combination of histidine and aspartate residues in the catalytic site, for which the HD domain and its superfamily are named[19].

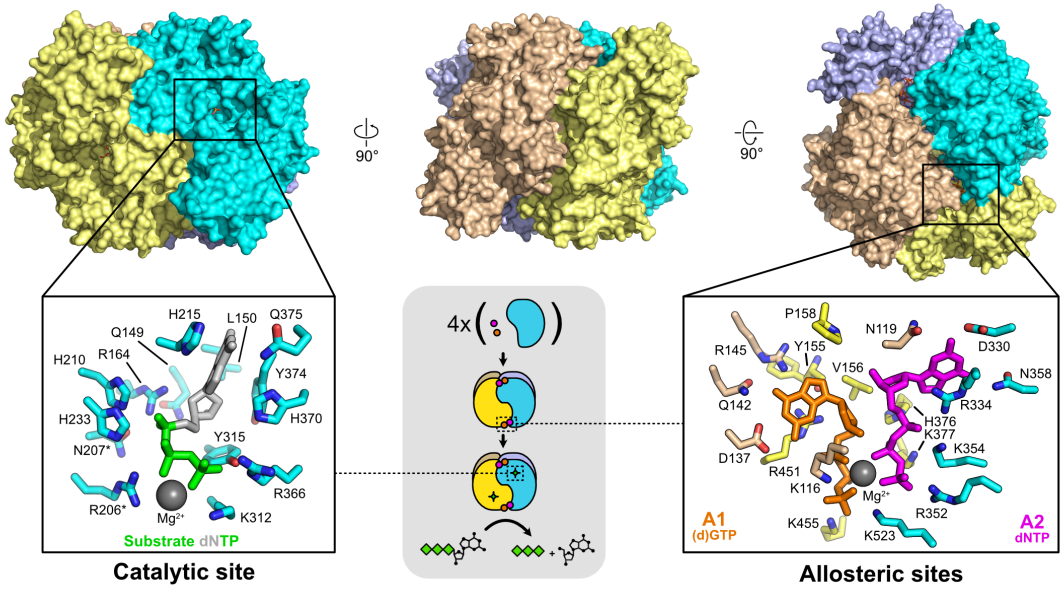

Figure 1-2. Quaternary structure and detailed dNTP-binding interactions of the human HD domain. Each subunit is colored separately. Two dimers (A/D and B/C, colored gold/wheat and cyan/lavender) come together upon binding GTP or dGTP at A1 (orange) and any dNTP at A2 (magenta) to become enzymatically active. Once assembled, catalytic site-associated triphosphohydrolase activity cleaves any dNTP into its component dN (light gray) and TP_i (lime green) products. Each nucleotide binding interaction is conserved within the other three subunits of the tetramer.

Typical dNTP concentrations can range from ~20-40 nM in differentiated, nondividing human macrophages to ~2-5 mM in activated human CD4+ T-cells according to a reverse transcriptase-based assay[20, 21]. Elegantly, the mechanism of SAMHD1 nucleotide-induced allosteric activation is fine tuned to parse dNTPase

activity between cell types due to a ~2-20 μM nucleotide activation threshold[22-24] essential for maintaining the nM dNTP concentrations in resting CD4+ T-cells and macrophages[20]. Once assembled, tetramer dissociation kinetics are partially dependent on dNTP concentration and which dNTP is bound to A2 and can allow for a long-lived state on the order of hours[23].

SAMHD1 is expressed in nearly all hematopoietic lineage cells as well as many “frontline” immunologically important tissues such as anogenital mucosa, among others[25]. Protein levels are highest in nondividing macrophages, dendritic cells, and resting CD4+ T cells while levels are lowest in activated CD4+ T cells where DNA polymerases require higher dNTP concentrations[25]. Type I interferon (IFN) stimulation induces expression in microglia, astrocytes, hepatocytes, monocytes, 293T cells, and HeLa cells, as reviewed by Coggins et al[26]. Interestingly, SAMHD1 expression is not enhanced by IFN-I in primary macrophages, dendritic cells, and CD4+ T cells due to high endogenous expression but a reduction in the fraction of phosphorylated SAMHD1 (pSAMHD1) is observed in those cell types[25, 27-30].

The presence of a nuclear localization signal (NLS)[31, 32] targets SAMHD1 to the nucleus, yet protein can be transiently localized or stably maintained in the cytosol depending on cell type and environment[28, 33-35]. Despite the necessity to regulate its dNTPase activity for DNA synthesis during cell cycle progression, SAMHD1 protein expression level is generally consistent throughout the cell cycle[36]. Regulation is therefore accomplished principally via nucleotide-induced allosteric activation and post-translational modification (PTM). SAMHD1 can be

highly decorated with various PTMs as noted in Figure 1-1. Some of these PTMs minorly modulate SAMHD1 activities. For example, K405 is located in the HD domain and acetylation increases dNTPase activity to achieve low dNTP pools and aid in the G1/S phase checkpoint transition[7]. N-Acetylglucosamine (GlcNAc) modifications on S93 of the SAM domain enhance tetramer stability and antiviral activity against Hepatitis B virus (HBV) in certain metabolic contexts[37]. Oxidizing cellular environments can lead to inhibitory oxidation on surface cysteine residues of the HD domain and attenuate dNTPase and antiviral activity[38, 39].

Other modifications take on a more major role in determining SAMHD1 function and largely localize to the HD domain and C-terminal regulatory domain. A very recent report asserts that SUMOylation of SAMHD1 is a major determinant of its antiviral activity in non-cycling cells[40]. Poly-SUMOylation accumulates on residues K469 and K622 upon proteasomal inhibition and mono-SUMOylation of residues K595 and K622 was observed. Interestingly, K595 is part of the ^{592}TPQK^{595} CDK recognition motif necessary for T592 phosphorylation (discussed below), and mono-SUMOylation at K595 appears to be necessary for HIV-1 antiviral activity in non-cycling cells regardless of phosphorylation state[40]. Residue K622 can also be ubiquitinated[41] in addition to being mono-/poly-SUMOylated, though factors governing regulatory crosstalk between these PTMs are unexplored.

Perhaps most intriguing of SAMHD1 modifications is phosphorylation of T592 by CyclinA2-CDK1/2, a kinase complex which recognizes C-terminal cyclin-binding ^{451}RxL^{453} and ^{620}LF^{621} motifs[27, 42, 43]. This motif is critical for tetramerization

since R451 stabilizes negative charge in A1-bound nucleotides, and alteration can negatively affect dNTPase activity, phosphorylation, and antiretroviral activity[43]. Phosphorylation is removed by cellular PP2A-B55α phosphatase during mitotic exit[36]. Phosphomimetic mutations such as T592D have been employed interchangeably to study the behavior of pSAMHD1 and its ambiguous functions. The effects of the T592D mutation were initially proposed to alter the ability of SAMHD1 to pack as a tetramer and thereby negatively regulate its dNTPase activity[42, 44], but other work refuted this model by arguing little change in k_{cat} or K_m versus WT caused by pSAMHD1[23, 45]. T592D has no significant effect on tetramer equilibrium or dNTPase activity of SAMHD1 *in vitro* and instead has been shown to influence tetramer dissociation kinetics or cellular localization.[24, 45, 46] This may help explain the inability of pSAMHD1 and T592D to restrict retroviruses due to the absence of a long-lived, catalytically active tetramer[24, 45]. Phosphorylation at this site is a key PTM exploited to modulate SAMHD1's function in many biological backgrounds as described below.

1.2 SAMHD1 functions in viral infection

As an enzyme with a critical function in maintaining one of the most fundamental building blocks in the cell, mutations can have a severe effect on cellular phenotype and protein function. The first hints of SAMHD1 as a major clinical target came with reports of mutations found in a subset of patients with Aicardi-Goutières syndrome (AGS), an uncommon genetic encephalopathy that phenotypically mimics congenital viral infection and systemic lupus erythematosus (SLE)[47, 48] A central characteristic of AGS includes aberrant Type I IFN responses

and upregulated IFN-stimulated genes (ISGs). This “hyper”-IFN response can be attributed to disrupted nucleic acid sensing/editing in the cell, and mutated genes include ADAR1, IFIH1, RNASEH2, TREX1, as well as the focus of this work - SAMHD1[47, 48].

SAMHD1 is perhaps most well-known for restricting replication of human immunodeficiency virus 1 (HIV-1) before the reverse transcription step, which completes in the nucleus[49]. Figure 1-3 shows the replication cycle of HIV-1. As a monomer or dimer SAMHD1 is capable of binding single stranded nucleic acids, and other work has demonstrated specific binding to the HIV-1 long terminal repeat (LTR) and interaction with integrated proviral DNA, where it inhibits LTR transcription[50-52]. SAMHD1 may also similarly inhibit the LINE-1 retrotransposon, which is known to be independent of dNTPase activity[53]. However, SAMHD1 has been shown to be a general viral restriction factor owing to its dNTPase activity. Indeed, depletion of the cellular dNTP fuel required for viral replication is sufficient to restrict a striking array of RNA and DNA viruses, including candidates from the families *Retroviridae*, *Papillomaviridae*, *Poxviridae*, *Picornaviridae*, and *Herpesviridae* among others[26]. Reported AGS mutants are incapable of restricting HIV-1 infection save for G209S, whose distinct retention of antiviral activity is separate from its aberrant IFN induction in AGS[54].

It is abundantly clear from the list above that SAMHD1 poses a substantial block on most viral replication and multiple viruses have co-evolved distinct mechanisms of antagonizing the restriction factor. Lentiviruses (e.g. HIV-1 and simian immunodeficiency viruses, SIV) are a genus within the family *Retroviridae*

and are uniquely able to infect both non-cycling and cycling cells[20]. SAMHD1 is actively dispensed in cases of SIV and HIV-2 infection with the accessory proteins Vpx or Vpr, which act as an adaptor to E3 ligase machinery ultimately resulting in ubiquitination and proteasomal degradation. Vpx-mediated antagonism, which occurs in the nucleus, is ablated upon mutating the NLS of SAMHD1 and results in cytoplasmic localization in which SAMHD1 retains antiviral and dNTPase activities[55]. HIV-1 conspicuously lacks an accessory protein that antagonizes SAMHD1[56]. Instead, HIV-1 reverse transcriptase (RT) has evolved to synthesize DNA even at low dNTP concentrations found in non-cycling cells like macrophages[57] and studies have suggested Vpx packaging with HIV-1 significantly enhances infection of non-cycling cells[58], an undesirable tradeoff of activating innate cytoplasmic DNA sensors because of elevated viral cDNA that can mount an IFN-1-mediated antiviral response[59].

Enteroviruses have been shown to thwart SAMHD1 restriction through upregulation of the host protein tripartite motif 21 (TRIM21), which is a component of innate immunity that directly binds and ubiquitinates SAMHD1 for degradation[41]. All *Herpesviridae* subfamilies encode a protein kinase which uses a unique mechanism among SAMHD1-sensitive viruses involving phosphorylation at T592 to inhibit dNTPase activity and relieve viral restriction[60]. The importance of the T592 regulatory cluster for antiviral activity is evidenced by strong positive selection of residues around T592 resulting from ancient evolutionary pressure from viruses, which was observed for SAMHD1 homologues from all mammals for which genetic data was available[61]. Even the SAMHD1 orthologue dGTPase is

directly targeted and inhibited by an accessory protein named Gp1.2 from the *Escherichia coli* bacteriophage T7[62]. It is clear that throughout evolution viruses targeting vastly different organisms must all grapple with HD domain activity to efficiently replicate, and poses an interesting scenario in which SAMHD1 dNTPase and antiviral activity must be continually adapted and honed while at the same time conserving functions that contribute to fundamental cellular maintenance.

While HIV-1 might be the most researched viral target of SAMHD1, precise mechanisms of HIV-1 restriction remain incomplete. As stated above, the suitably high dNTP concentrations found in cycling or activated cells render them more HIV-permissive. Although pT592, a biomarker for cycling cells, and phosphomimetic mutants cannot restrict HIV-1, they still retain dNTPase activity[24, 63]. Thus, pT592 is a major determinant for differential HIV-1 restriction competency between cell types and activation states[63]. In non-cycling cells, SAMHD1 has been shown to block HIV-1 infection through kinetic suppression of reverse transcription, a replicative step requiring consumption of cellular dNTPs for proviral synthesis[64, 65]. Vpx-mediated degradation of SAMHD1 in macrophages results in a transient spike in dNTP levels which provides evidence that SAMHD1 is indeed catalytically active[65, 66]. Other work demonstrates that the dNTP levels found in non-cycling HIV-restrictive cells are however 2-4 orders of magnitude below the threshold of SAMHD1 allosteric activation, and the authors raise questions about the relevance of dNTPase activity of SAMHD1 for the primary mechanism of HIV restriction in non-cycling cells[23]. Only recently has K595 SUMOylation been shown to be yet

another essential component for restriction (in non-cycling cells), irrespective of T592 phosphorylation status[40].

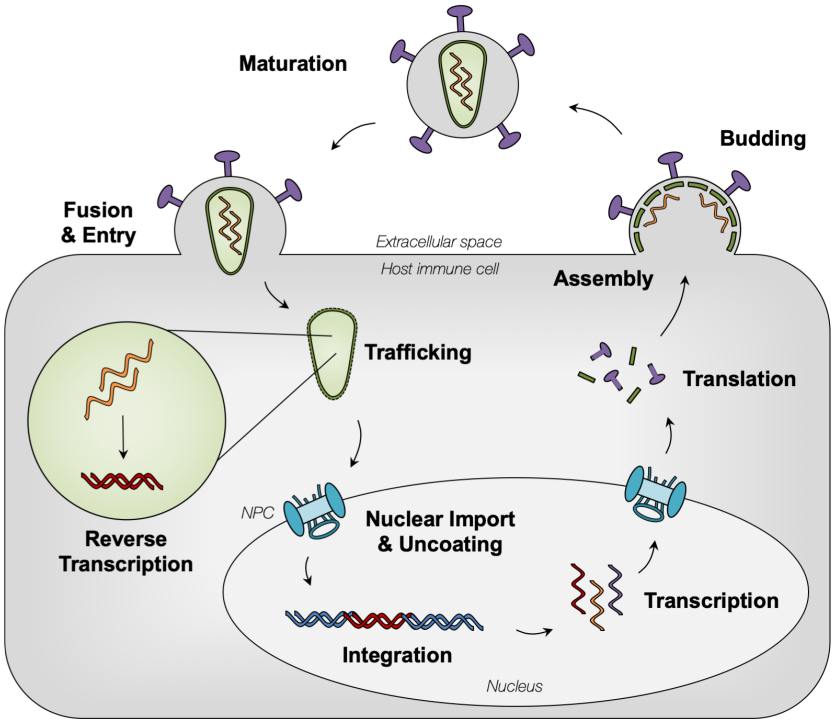

Figure 1-3. Basic steps of the HIV-1 replication cycle. Docking and membrane fusion of the virus is initiated by Env protein (purple). The viral capsid (green cone) is deposited and trafficked to the nucleus as reverse transcription begins. Mostly intact capsid is imported through NPCs where it disassembles. Reverse transcription completes in the nucleus where the viral cDNA integrates within the host genome. Viral genes are then expressed to assemble immature progeny virions at the plasma membrane. Full maturation of the virion occurs following budding.

How SAMHD1 is activated given low dNTP concentrations in non-cycling cells and then accomplishes restriction under the combinatorial influence of phosphorylation and SUMOylation remains an open question. Furthermore, how a phosphomimetic mutation is unable to restrict HIV-1 infection given a purely kinetic effect on tetramer formation is unclear. Some authors have proposed the possibility of a distinct mechanism of SAMHD1 activation in non-cycling cells redundant and/or independent of nucleotide-induced allostery[23]. This is supported

by data of a SAMHD1 mutant (Y146S/Y154S) that is tetramer-defective *in vitro* yet still able to deplete dNTPs and restrict HIV-1 in macrophages[23, 31, 45]. Furthermore, the ability of pSAMHD1 and phosphomimetics to restrict HIV-1 and deplete cellular dNTPs does not always correlate with *in vitro* dNTPase activity and tetramer formation of purified recombinant protein, which probably indicates an undiscovered cellular factor contributing to SAMHD1 activation through interaction or post-translational modification[23, 45].

Progression of HIV-1 infection is characterized by a shift in preference from permissive cycling cells to establish early infection, namely activated CD4+ T-cells, to more restrictive non-cycling cells in chronic infection, especially macrophages which are thought to be a critical reservoir of latent HIV-1 as CD4+ T-cells are depleted; this subject is reviewed by Rodrigues et al[67] and the authors highlight that not all macrophages are functionally equivalent, often with specific cellular functions due to tailored transcriptional programs depending on tissue residency. Macrophages also have exclusive cell biological processes during HIV-1 infection since the virus buds into a distinct endosomal structure called the virus-containing compartment (VCC), not out of the plasma membrane like in other cell types[68]. One novel report even suggests HIV-1 infection of primary macrophages, but not CD4+ T-cells, may be stratified by sex. The authors demonstrated that female-derived macrophages were less susceptible to HIV-1 infection and exhibited lower pT592 levels, although this preliminary study needs corroboration with larger sample sizes[69]. Thus, given the intrinsic relationship between SAMHD1 activity and cell type/activation state, the potential for major differences in regulation are

evident and warrant further detailed analysis of regulatory mechanisms on structural, biochemical, and cell biological levels.

1.3 SAMHD1 is implicated in cancer and innate immune suppression

Mutants found in diverse cancers also reveal the overarching importance of SAMHD1's role in cellular homeostasis and genetic pathology. A hallmark of cancerous cells is an aggressive replicative state in which cells are actively consuming dNTPs to fuel DNA synthesis for cell division. It is no surprise then that increased dNTP levels (by roughly 6-11 fold[70]) are a biochemical hallmark of cancerous cells, as reviewed by Amie et al[1], and atypical dNTP levels correlate with tumor formation and genomic instability[71, 72]. SAMHD1 has been implicated in DNA repair processes that also work toward staving off tumorigenesis[73, 74], which includes ensuring telomere stability[75].

In order to maintain a cancerous cellular state SAMHD1 must be deactivated, and SAMHD1 is indeed shown to be down regulated or mutated in cases of chronic lymphocytic leukemia (CLL)[76, 77], lung cancer[78], cutaneous T-cell lymphoma (CTCL)[79, 80], acute myeloid leukemia (AML)[81], colon cancer[82], and T-cell prolymphocytic leukemia[83] for example. Searching the SAMHD1 gene in the Catalogue Of Somatic Mutations In Cancer (COSMIC), which curates human somatic cancer data from publications all over the world, reveals 633 unique cancer-associated mutations across 142 publications as of the time of this writing in August 2021. COSMIC shows accumulated mutations distributed throughout the entire protein length much like for AGS mutations[47, 54]. It is important to investigate the mechanism of SAMHD1 dysfunction in these cancerous backgrounds which

can shed light on the protein's endogenous function and regulation as well as guide higher efficacy cancer therapies.

Immunosuppression is another recently described pillar of SAMHD1 function. In cases of dsDNA repair via homologous recombination (HR), SAMHD1 must interact with the protein CtIP (C-terminal binding protein 1 interacting protein) to facilitate DNA end resection[73]. In addition, SAMHD1 aids a separate nuclease, Mre11, in processing nascent DNA at stalled replication forks during DNA replication which consequently inhibits IFN induction; taken together, these functions are dNTPase-independent and help prevent the host from mounting an autoimmune response against self-nucleic acids[74]. SAMHD1 also interacts with and negatively regulates immunologically important transcription factors NF-κB and IRF7 in addition to interactions with proteins in both immune pathways. By inhibiting phosphorylation of the NF-κB inhibitory protein (IκBα) SAMHD1 is able to prevent nuclear translocation and therefore activation of the NF-κB signaling axis; similarly, SAMHD1 is able to reduce inhibitor-κB kinase ε (IKKε)-mediated phosphorylation of IRF7, thus preventing IRF7 dimerization and nuclear entry for ISG activation[34, 84]. NF-κB suppression appears dependent on SAMHD1 dNTPase activity in monocyte-derived cells, yet independent of dNTPase activity in dividing 293T cells[34] which further establishes a key missing component in understanding activation mechanisms of SAMHD1. Despite being a general antiviral factor (and as with any SAMHD1 function) there are caveats – some viruses actually exploit SAMHD1 immunosuppressive activity in certain contexts. SAMHD1 dNTPase activity is important to suppress NF-κB activation during infection by human

cytomegalovirus (a DNA virus)[85] much like by Sendai virus (SeV; an RNA virus)[34]. A more recent corollary in SARS-CoV-2 shows genetic loss of SAMHD1, not just a lack of dNTPase activity, activates cellular innate immunity which thereby suppresses SARS-CoV-2 replication[86]. SAMHD1 can also support the infection of Zika and Chikungunya viruses, although through uninvestigated mechanisms[87].

Taken together, this highlights the incredibly dynamic role of SAMHD1 at the crossroads of cellular metabolism, viral infection, cancer, and immunity. Characterizing mutants that uncouple SAMHD1's often-intertwined functions will aid in finding missing determinants in SAMHD1 activation and regulation in these contexts. As a highly allosteric molecule, not much information is known regarding structural differences or altered dynamics induced by PTMs that may help explain its functional differences. Furthermore, no structural data exist in the context of the full-length human protein. These factors combined serve as the motivation for this study.

1.4 Aims and Scope

Ultimately this dissertation is a culmination of many projects centered around interactions in the molecular arms race between viruses and human immunity, otherwise termed the host-virus interface. Some of these initially promising projects did not result in fleshed out biological stories yet are detailed as appendices below to document significant partial or negative data. These topics include stability and binding interactions of the HIV-1 capsid in addition to cryo-EM methods development to solve a critical bottleneck in obtaining high resolution cryo-EM structures. Here the primary focus however is on investigating structural

mechanisms of the protein SAMHD1 function and dysfunction in various contexts. The work presented here is aimed at characterizing clinically important SAMHD1 mutants and addresses basic structural regulation of SAMHD1, with some discussion of processing data from challenging crystals.

Nucleotide-dependent assembly and allosteric activation of SAMHD1 is likely the most well-characterized of the protein's many functions, but that represents one small facet of SAMHD1 biology. Because of all the multifaceted functions detailed above, SAMHD1 is also interesting from a virology, cancer biology, and immunology perspective. Characterizing mutants of SAMHD1 is valuable in isolating and understanding determinants for these functions. Despite its implication in various cancers, not much literature is devoted to understanding the mechanisms of SAMHD1 dysfunction therein.

Furthermore, zero to little structural information exists for the full-length or for post-translationally modified protein, which is important to help solve the puzzle of regulation of SAMHD1 activity. All current structural models of human SAMHD1 lack the N-terminal SAM domain since it is largely dispensable for dNTPase activity and retroviral restriction ability[32] and the only full-length structure is the mouse orthologue[88]. Previous HD domain crystal structures are unable to resolve density beyond residue ~580 for T592E[44] or *in vitro* phosphorylated SAMHD1[24] despite ordered density being observed up to position 599 in non-modified catalytically-inactive $SAMHD1_{RN}$ (H206R/D207N)[17]. Combined, these data indicate that the N-terminal SAM domain and C-terminal regulatory domain are labile. Along with high symmetry and mass (~290 kDa) of the tetramer, this makes SAMHD1 a suitable

candidate for single particle cryo-EM to parse the distribution of dynamic states and resolve structural differences therein. Understanding structural regulation of SAMHD1 beyond basic allosteric assembly will provide critical insight into its overall behavior. The following studies constitute the first structural characterization of novel cancer-causing mutations in the SAMHD1 catalytic core using X-ray crystallography and in addition the first cryo-EM structures of human SAMHD1 are presented.

2. Structural insight into cancer-associated SAMHD1 mutations

2.1 Introduction

Here, selected mutations from a panel of recent reports detailing SAMHD1 mutants associated with CLL and colon cancer are studied to uncover their mechanism of dysfunction. These mutations are located in key nucleotide binding sites of SAMHD1 (Figure 1-2). Selected CLL mutants include R145Q, Y155C, P158S, and R451C which are all located in the A1/A2 binding sites. Two reported R366 mutants localize to the catalytic site, R366C and R366H. R366C has been reported in CLL[89] whereas R366H has been found in colon cancer[82, 90]. Two other CLL mutants, I201N and L244F[76], do not reside in positions normally contacting bound nucleotides.

We ultimately focused on these R366 mutations, which are not AGS-associated and intriguingly can result in different types of cancer depending on the mutant side chain identity. As SAMHD1 has a wide range of reported biological activities, the contribution of each to overall cancer pathogenesis must be evaluated. Our structural work complemented a functional analysis of these mutants as part of a collaboration with co-first author Nicole Bowen from the lab of Baek Kim, who led the work[91]. Selected data collected by other authors from this manuscript are presented here with attribution to help detail the complete analysis of these mutants. We examine stability of the mutant proteins, their ability to tetramerize, and their dNTPase activity. We also use X-ray crystallography to determine the structure of the R366 mutants and further investigate whether the mutants retain dNTPase-independent functions that may be implicated in cancer

cell phenotypes. This study suggests that impaired dNTPase activity, not dNTPase-independent functions, is a mechanism by which SAMHD1 R366 mutants can contribute to leukemia and colon cancer proliferation, while many other mutations simply result in reduced SAMHD1 protein stability.

2.2 Results

2.2.1 Stability and tetramerization of cancer-associated SAMHD1 mutants

Mutant screening began with Western blotting to analyze protein levels following transfection of HEK293T cells. The majority of sampled mutations – R145Q, Y155C, P158S, I201N, L244F, and R451C – markedly reduced SAMHD1 protein levels (Figure 2-1a). Two notable exceptions were R366C and R366H, which were both comparable to WT SAMHD1 levels. We tested the thermostability of these two mutants after incubation with dGTP using a fluorescence-based unfolding assay in which hydrophobic protein cores that are exposed upon thermal denaturation interact with a fluorescent dye and increase its fluorescence emission[92]. The temperature at which half-maximal unfolding is observed for the R366 mutants (R366C: 62.0 ± 0.4 ˚C and R366H: 62.3 ± 0.1 ˚C) were modestly lower than WT (63.5 ± 0.1 ˚C) yet still higher than the Y155C mutant (60.2 ± 0.3 ˚C) which served as a representative of an unstable mutant since trace protein levels were detected via western blot (Figure 2-1a, b). The denaturation profiles exhibit a minor unfolding event at ~40 ˚C for each mutant that is not observed for WT. Formaldehyde crosslinking was then used to analyze the ability of these mutants to tetramerize and confirmed that R366 mutants, but not the Y155C mutant, are able to form tetramers comparable to WT SAMHD1 when supplied

with dGTP (Figure 2-1c). As Vpx-mediated antagonism in HIV-2 and SIV infection is dependent on the tertiary fold of SAMHD1[93], a Vpx degradation assay showed a similar ability of Vpx to dispatch WT, R366C, and R366H SAMHD1 thereby indicating a generally conserved fold between WT and R366 mutants (Figure 2-1d).

2.2.2 R366 mutants have impaired dNTPase activity and cannot restrict HIV-1

Next, we employed a thin layer chromatography (TLC)-based assay and α^{32}P-dGTP to monitor dNTP triphosphohydrolase activity of each mutant (Figure 2-2a). Of this panel of mutants, only R366H maintained partial activity and accumulated ~25% of the radiolabeled triphosphate product compared to WT. Interestingly this partial activity was observed with hydrolysis of dGTP alone, not with other dNTPs (Figure 2-2b). To complement *in vitro* experiments R366 mutants were expressed in differentiated U937 cells, which lack endogenous SAMHD1 expression, and antiviral competency and dNTP levels were compared with WT- and D311A-expressing cells. The D311A mutation abrogates metal ion binding in the catalytic site and renders SAMHD1 deficient in HIV-1 restriction and dNTPase capabilities[94] thus serving as a negative control. Indeed, expression of D311A, R366C, and R366H all failed to suppress HIV-1 infection in differentiated non-cycling U937 cells, unlike WT SAMHD1 (Figure 2-3a).

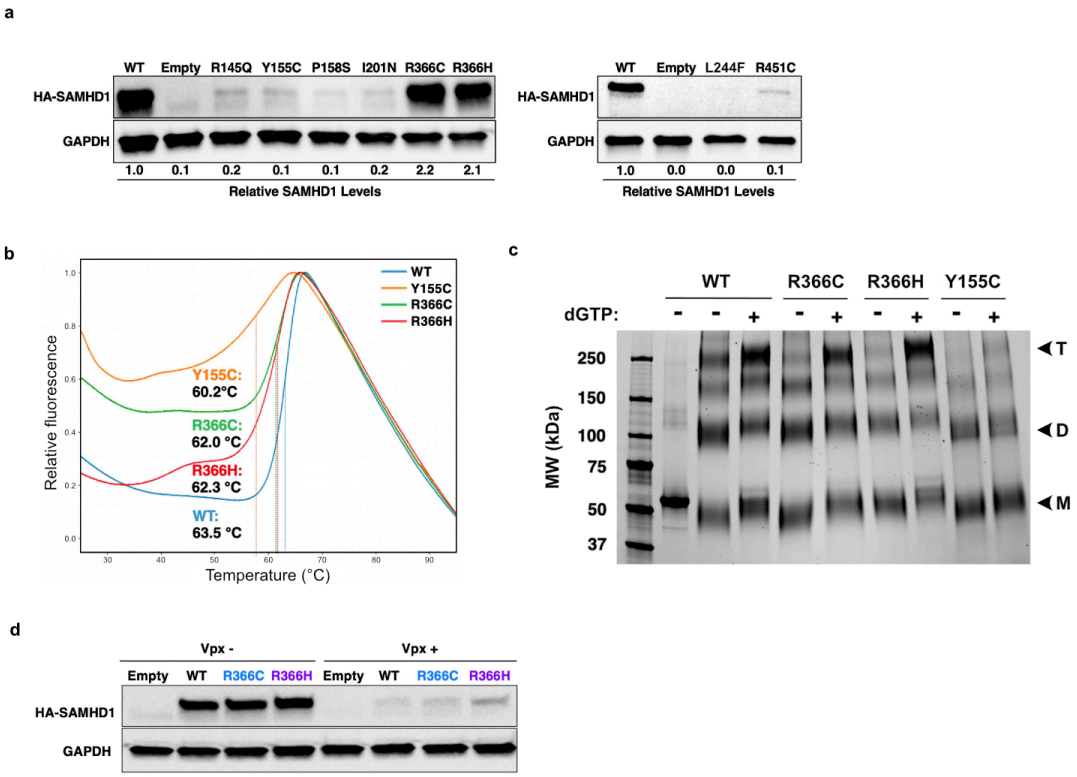

Figure 2-1. Characterization of protein stability, tetramerization capability, and Vpx degradation of selected cancer-associated SAMHD1 mutants. A) The SAMHD1 protein levels in HEK293T cells transfected with an equal amount of plasmids expressing HA-tagged SAMHD1 proteins. Empty: Backbone plasmids, pKH3-3xHA (left) and pLVX-IRES-mCherry (right). The relative SAMHD1 protein levels were normalized by GAPDH protein level and the ratios between wild type and mutant SAMHD1 protein levels were calculated. B) Thermal shift assay of wild type and mutant HD proteins was conducted after preincubation with SYPRO Orange dye. The melting temperature (Tm) of each protein was calculated as described in Methods. Tm: WT = 63.5 ± 0.1°C, Y155C = 60.2 ± 0.3°C, R366C = 62.0 ± 0.4°C, R366H = 62.3 ± 0.1°C. C) Tetramerization of wild type and mutant HD domain proteins was analyzed by SDS-PAGE after formaldehyde crosslinking in the presence (+) and absence (-) of 2 mM dGTP. M: Monomer, D: Dimer, T: Tetramer. MW: Molecular weight. C: No formaldehyde control. D) Vpx-mediated proteasomal degradation of wild type and mutant SAMHD1 proteins in cells was monitored. 293T cells were co-transfected with SAMHD1 expressing plasmids and Vpx expressing (Vpx+) or non-expressing (Vpx-) plasmid. SAMHD1 protein levels in the transfected cells were determined by immunoblot with anti-SAMHD1 antibody.

Cellular dNTP concentrations were generally consistent with biochemical data and R366C/H-expressing cells had statistically significant elevated dNTP levels versus WT (Figure 2-3b). Increased concentrations of dATP, dCTP, and dTTP are observed for R366 mutants relative to the D311A construct. Curiously, activity of R366 mutants exhibits a unique effect for dGTP similar to *in vitro* data since R366 mutants have depressed dGTP levels compared to D311A, a phenomenon not observed for other dNTPs. Lastly, the inability to deplete dNTPs was confirmed in patient-derived cells. Virus-like particle (VLP)-mediated treatment with Vpx in activated primary CD4+ T-cells disposed of ~30% of cellular SAMHD1 and resulted in significant increases in concentrations of all dNTPs (Figure 2-3c). The magnitude of the relief in SAMHD1-mediated suppression for dATP and dGTP is approximately double the effect observed for dCTP and dTTP.

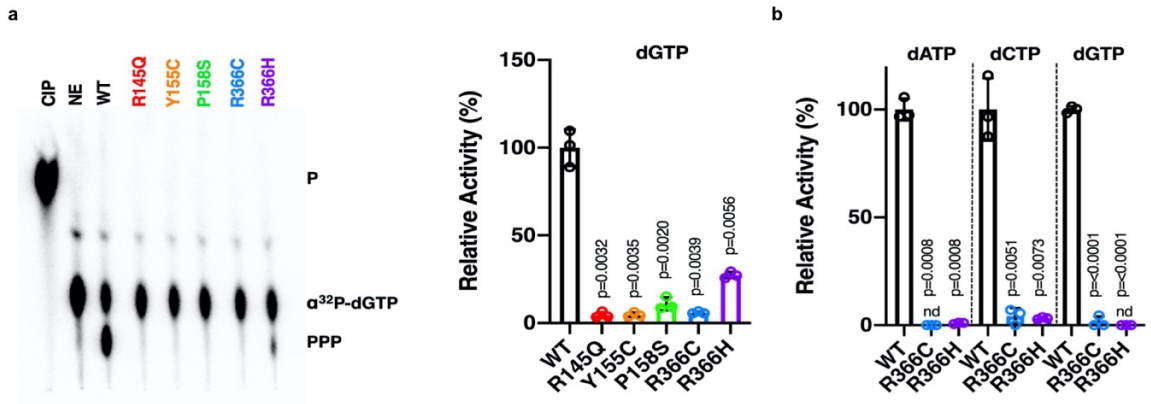

Figure 2-2. A) *In vitro* dNTPase activity of recombinant SAMHD1$_{113-626}$ using TLC to detect α^{32}P-dGTP hydrolysis. Relative dGTPase activity of wild type and mutant SAMHD1 proteins (HD domain) was calculated by dividing the triphosphate product by the lane total and normalizing to wild type dGTPase activity. B) The relative dNTPase activities of wild type, R336C and R366H proteins were determined for dATP, dCTP, and dTTP. Data are the mean of three replicates and error bars reflect standard deviation from the mean. P values were determined using two-tailed, unpaired Welch's t-test with WT, wild type. CIP: calf-intestinal phosphatase control. NE: No enzyme negative control. P: Monophosphate. PPP: Triphosphate. nd: not detected. Figure and legend were adapted from manuscript[91].

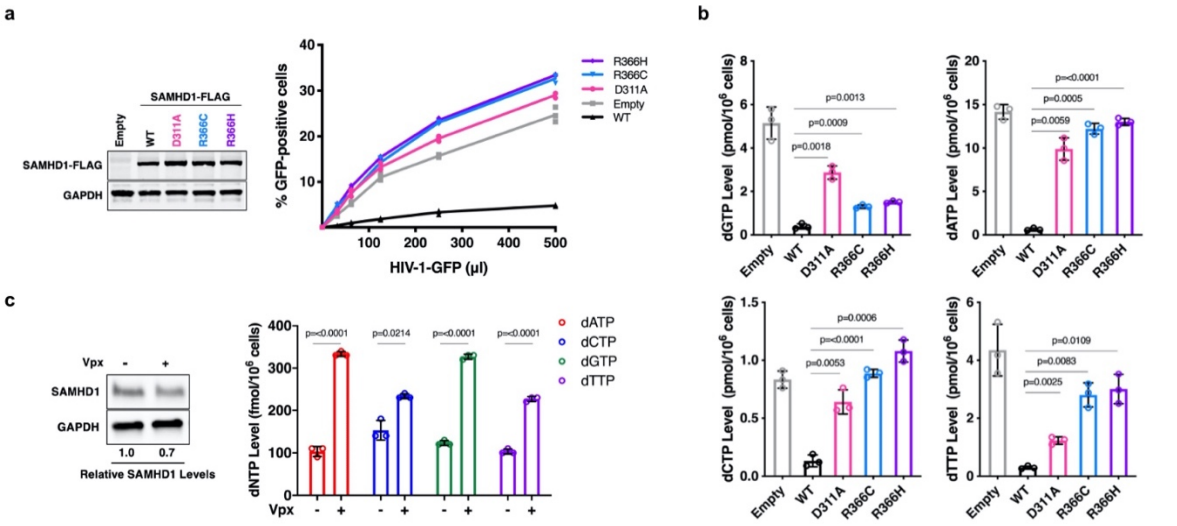

Figure 2-3. HIV-1 restriction competency and cellular dNTP effects. A) Left, U937 cells were transduced with lentiviral vector expressing FLAG-tagged wild type, D311A inactive mutant, R366C, and R366H proteins. The expression level of each SAMHD1 protein was determined by western blot with anti-FLAG antibody and normalized with GAPDH. Right, the transduced U937 cells were differentiated to nondividing stage, and transduced with eGFP expressing HIV-1 vector. Transduction efficiency using different quantities of HIV-1 vector was determined using flow cytometry. B) Intracellular dNTP levels in differentiated U937 cells expressing wild-type and mutant SAMHD1 proteins were determined by RT-based dNTP assay. Data are the mean of three replicates and error bars reflect standard deviation from the mean. P values were determined using two-tailed, unpaired Welch's t-test to WT knock-in cells. C) Human primary CD4+ T cells were isolated from 3 healthy donors, activated by PHA and IL-2 for 5 days, and treated with VLP Vpx (-) or VLP Vpx (+) for 24 hours. Cellular SAMHD1 protein levels were determined by immunoblot using anti-SAMHD1 antibody and anti-GAPDH antibody as a loading control. The relative SAMHD1 protein levels were normalized by GAPDH protein level and the ratios between wild type and mutant SAMHD1 protein levels were calculated. Intracellular dNTP levels were determined by RT-based dNTP assay. Data are the mean of three replicates and error bars reflect standard deviation from the mean. P values were determined using two-tailed, unpaired Welch's t-test to VLP Vpx (-) condition. Figure and legend adapted from manuscript[91].

2.2.3 Crystallization and structure determination of R366 mutants

Full-length human SAMHD1 has thus far been unable to be crystallized in contrast to the catalytic core, which is readily crystallized. For both R366C and R366H mutants, we therefore crystallized the HD domain and C-terminal regulatory region of SAMHD1 (residues 113-626) in the presence of dGTP, which is capable of binding both A1 and A2 as well as the catalytic site of SAMHD1[10]. In the WT catalytic site the guanidinium group of R366 neutralizes the negative charge of the substrate dNTP γ-phosphate and interacts with D506 to stabilize the bound nucleotide (Figure 2-4, cyan). The catalytically inactive mutant $SAMHD1_{RN}$ was used to prevent hydrolysis but retain binding capacity of dGTP during crystallization.

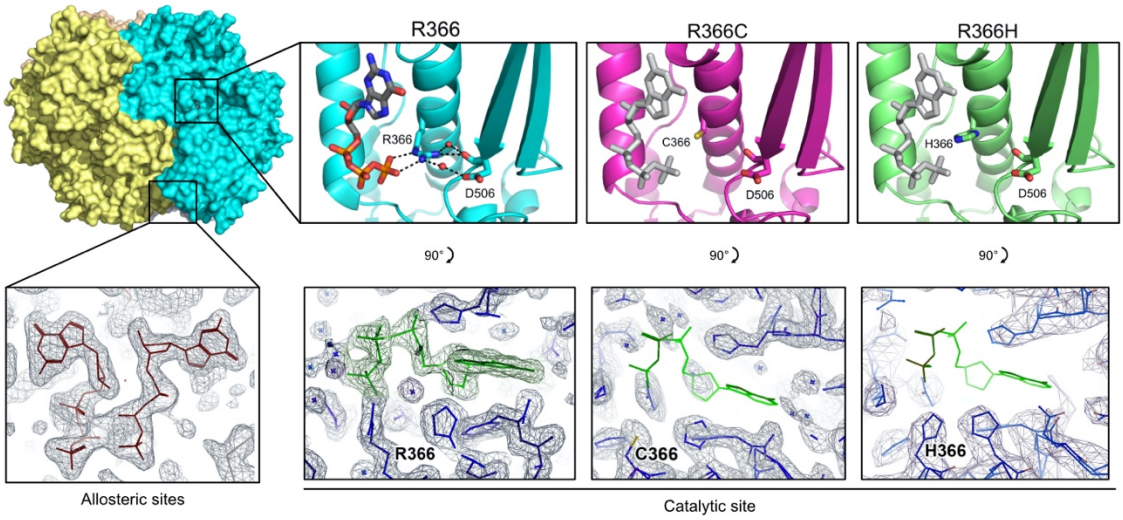

Figure 2-4. R366C/H mutations abrogate catalytic nucleotide binding. Overall structure of the SAMHD1 tetramer in surface representation (top left). The upper row of insets shows selected interactions of R366 and the catalytic nucleotide in $SAMHD1_{RN}$ (PDB code 4BZB). H-bonds and salt bridges are shown as dashed lines. R366C (magenta cartoon) or R366H (green cartoon) mutation leads to the disruption of the interactions and the loss of nucleotide binding at the catalytic site. The catalytic nucleotides (gray) are modeled in based on their theoretical positions in $SAMHD1_{RN}$. Portions of the structure have been hidden for clarity. The bottom row shows representative 2Fo-Fc density at the allosteric sites (for all structures) and the catalytic site (for each individual structure).

R366C crystals were fewer and formed on the span of weeks as singular rectangular prisms with a deep imperfection (Figure 2-5a). These crystals were difficult to replicate in further optimization attempts yet were stable on the order of weeks and diffracted to high resolution at 1.9 Å. R366H crystals formed rapidly overnight and fully matured within a week. Streak seeding was used to improve thick, flat faceted prisms which were generally clustered around a central nucleus with multiple adsorbed lattices (Figure 2-6a). Indeed, the diffraction data collected for R366H were twinned and contained data from multiple crystals which limited the final resolution to 3.6 Å. Statistics for data collection are summarized in Table 2-1.

$SAMHD1_{RN}$ R366C/H mutants both retained a tetrameric architecture in the crystal structures and showed no significant deviations from a previously solved dGTP-bound $SAMHD1_{RN}$ tetramer ($RMSD_{R366C}$: 0.36 Å and $RMSD_{R366H}$: 0.39 Å when compared with PDB 4BZB). Importantly, no density was observed in the catalytic pocket of either mutant, unlike the allosteric sites each with appropriate density for dGTP (Figure 2-4). Of note, additional amorphous density is observed close to the R366C side chain upon map sharpening (Figure 2-5b), consistent with an earlier 2.7 Å data set showing similar features (data not shown).

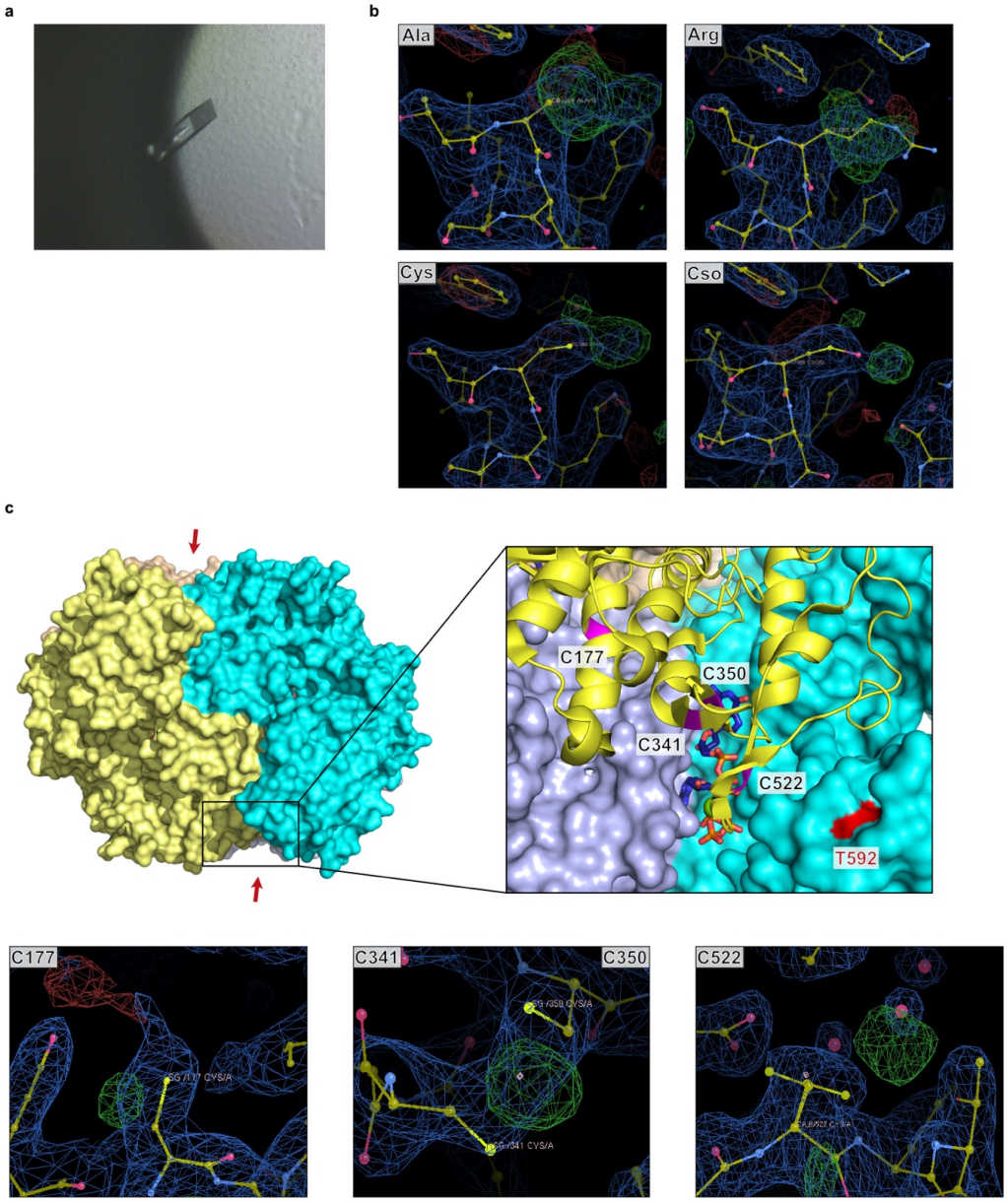

Figure 2-5. R366C crystals show abnormal electron density features surrounding cysteine side chains. A) Image of the R366C crystal. B) Representative 2Fo-Fc electron density (blue) and Fo-Fc density (green) of residue 366 modeled as various side chains to explain extra density including: Ala (alanine), to explore unbiased side chain features; Lys (lysine), to evaluate if the proper mutation is present; Cys (cysteine), the expected side chain identity; and Cso (S-hydroxycysteine), an oxidized form of cysteine. C) Location of other surface-exposed cysteines (magenta) with observed amorphous density are marked noting proximity to T592 (red) of the adjacent monomer and at the tetramer interface. Red arrows mark locations of T592.

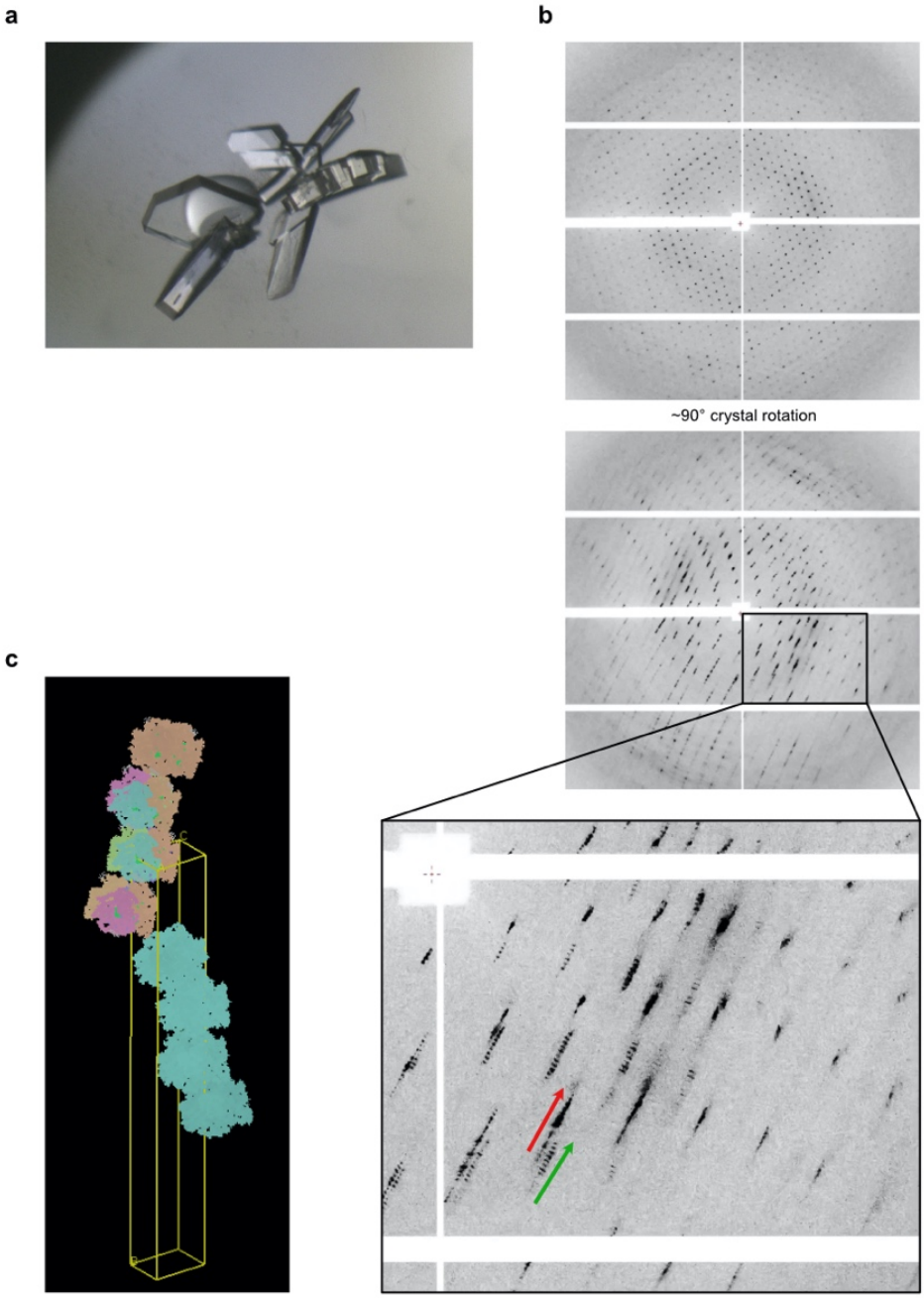

Figure 2-6. R366H crystal packing and other pathology gives rise to difficult diffraction data. A) Image of the R366H crystal. B) Diffraction image pairs taken roughly 90˚ orthogonal to each other. Data were anisotropic, mosaic, twinned, and contained reflections from multiple lattices. Green arrow indicates indexed set of reflections. Red arrow indicates contaminating lattice reflections. C) Packing of four tetramers in the asymmetric unit yielded an uncharacteristically large unit cell which contributes to the extremely fine spacing between reflections observed in panel B, lower. The discrete reflections in panel B, upper are columns of stacked reflections caused by the long *b* axis oriented parallel with the incident x-ray beam.

Table 2-1. Data collection and refinement statistics for R366 mutants.

	R366C	R366H
Data collection		
Wavelength (Å)	0.97918	0.97918
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, b, c (Å)	80.9, 140.1, 97.2	83.7, 573.5, 100.5
α, β, γ (˚)	90.0, 114.2, 90.0	90.0, 114.7, 90.0
No.molecules/ asymmetric unit	4	16
Resolution (Å)	50.0-1.9 (1.93-1.90)	50.0-3.60 (3.66-3.60)
R_{merge}	0.071 (>1)	0.143 (>1)
Mean $I / \sigma I$	17.3 (1.0)	13.8 (1.6)
$CC_{1/2}$	0.999 (0.297)	0.996 (0.690)
Completeness (%)	99.1 (98.8)	85.6 (83.1)
Redundancy	3.3 (3.3)	5.1 (4.8)
Unique reflections	153,730 (7,622)	84,897 (4,131)
Refinement		
No. nonhydrogen atoms	16,683	64,032
R_{work}/R_{free}	0.173/0.207	0.227/0.259
Mean *B*-factor (Å^2)	31	144
R.m.s.d.		
Bond lengths (Å)	0.012	0.008
Bond angles (˚)	1.7	1.4
Ramachandran		
Favored (%)	98.64	97.84
Allowed (%)	1.26	2.11
Outliers (%)	0.10	0.05

Parentheses indicate highest resolution shell.

2.2.4 SAMHD1 R366 mutants maintain other endogenous functions

It is important to evaluate the ability of SAMHD1 cancer-associated mutants to carry out other dNTPase-independent functions and ask whether these may contribute to cancer progression. The Kim lab completed a wide functional screen, and selected data are shown above. Further experiments (data not shown[91]) confirmed unaltered interactions between R366 mutants and Cyclin A2 via co-immunoprecipitation (co-IP), as well as DNA repair protein CtIP. Both mutants also had comparable HR repair activity to WT SAMHD1 in an RFP gene repair reporter

assay. Both R366 mutants also efficiently suppressed expression of a Luciferase (Luc) reporter gene under control of either the HIV-1 LTR or an IFN-stimulated response element (ISRE) similar to WT levels in 293T cells. Notably however, R366 mutants displayed a modest reduction in affinity for ssDNA and ssRNA oligonucleotides versus WT SAMHD1 using fluorescence polarization.

2.3 Discussion

Abnormally high dNTP levels are a biochemical hallmark of cancerous cells due to aggressive and uncontrolled cell proliferation which requires ample dNTP fuel[70]. This study shows that mutation of R366 causes impaired dNTPase activity and concomitant increase in cellular dNTP pools and can be mechanistically linked to CLL and colon cancer, independent of other biological functions of SAMHD1. In the majority of cases, cancer-associated mutant protein levels were generally $\lesssim$ 10% of WT SAMHD1. Similarly, previously described CLL mutations have been shown to cause reduced SAMHD1 protein levels and have been linked to cancer pathogenesis[76]. The R366 mutants are notable exceptions that retained cellular stability in addition to the capacity to tetramerize as evidenced by formaldehyde crosslinking and crystal structure determination discussed below (Figures 2-1c and 2-4). To our knowledge, no other disease-associated SAMHD1 mutant has been reported to retain tetramerization capability.

Despite the presence of an assembled tetramer, severely attenuated dNTPase activity is observed for R366H with near complete ablation of dNTPase activity for R366C (Figure 2-2). Moreover, dGTP-activated R366H partial activity is specifically biased towards hydrolyzing dGTP over other dNTPs (Figure 2-2a),

an effect also originally observed using GTP-activated R366H *in vitro*[82]. Both R366 mutants drive a decrease in dGTP levels relative to other dNTPs in transfected cells and treatment with Vpx relieves suppression of purine dNTP levels by WT SAMHD1 more significantly than for pyrimidines (Figure 2-3). Combined, these data show a biased sensitivity of SAMHD1 to dGTP which is interesting considering guanine nucleotide triphosphates are solely capable of binding A1 and priming SAMHD1 for allosteric activation[14, 22].

The $SAMHD1_{RN}$ catalytic core was then crystallized in the presence of dGTP, a single nucleotide sufficient for allosteric activation and catalytic substrate binding which also appeared to be preferred by R366H. Despite dGTP density at both allosteric sites and packing into a tetramer, no significant density was observed at the catalytic sites (Figure 2-4). This confirms that mutation of the arginine side chain at position 366 renders $SAMHD1_{RN}$ deficient in stably binding dGTP at the catalytic site and is consistent with R366 mutants' impaired dNTPase activity and concomitant increase in cellular dNTP pools.

Data processing presented challenging cases. R366C diffraction data showed dubious density directly adjacent to the mutated cysteine side chain of interest (Figure 2-5). Water was initially built into some of these blobs, but two scenarios are more likely. First, there is evidence of oxidation of other solvent-accessible cysteine side chains. Additional Fo-Fc difference density is also observed for C177 and C522 combined with a disulfide bridge between C341-C350. Taken together, it is clear $SAMHD1_{RN}$ R366C was in an oxidized state despite the presence of 0.5 mM TCEP in the protein buffer. SAMHD1 is reportedly

redox-regulated since oxidation of C341, C350, and C522 contribute to decreased tetramerization and dNTPase activity[38, 39], which may contribute to its mechanism of dysfunction or downregulation in CLL. It is also interesting to note that the cluster of oxidized cysteines lie immediately adjacent to the tetramer interface and T592 of an opposing dimer, as well as the theoretical region the SAM domain can sample. Alternatively, it is clear through crystallization and crosslinking data that R366 mutations do not inhibit tetramerization. Theoretically the catalytic pocket could therefore retain partial or transient binding capacity of substrate nucleotide causing weakly averaged partial density for dGTP to appear in the map.

R366H, purified identically to R366C, formed crystals that exhibited no amorphous density or evidence of oxidation. Despite crystallizing only 0.8 pH units higher in the otherwise same condition as R366C, the R366H crystal exhibited novel packing with four tetramers in the asymmetric unit. This packing generated a 573.5 Å axis and, along with another lattice adsorbed to the primary crystal, resulted in major artifacts in the diffraction data. The incredibly long *b* axis caused overlapping reflections in reciprocal space and drastically reduced signal:noise. Streaky, off-angle contaminant reflections are caused by an imperfect adsorbed crystal and even further complicated indexing which required parsing the more discrete set of reflections (Figure 2-6b, red arrow). Lastly the indexed diffraction data were twinned and altogether this limited resolution to 3.6 Å, yet still allowed us to determine whether nucleotide ligands were present. Uniquely, N ε2 of the mutant R366H side chain sits close to the γ-phosphate position in a modeled dGTP-bound mutant structure and is therefore theoretically capable of an

interaction, which could explain a higher relative R366H activity with dGTP versus other cancerous mutants and may contribute to the difference in cancer manifestation based on side chain identity.

Both R366 mutants efficiently tetramerize and theoretically could retain other oligomer-dependent, or simply dNTPase-independent, functions. In the majority of our assays not probing dNTPase activity or HIV-1 restriction, R366 mutant effects on other WT SAMHD1 functions were negligible, as observed for: degradation by Vpx; co-IP with CtIP and HR-mediated DNA repair; co-IP with Cyclin A2; and suppression of HIV LTR-driven and ISRE-mediated gene expression. Some differences are noted however in the reduced ability of the recombinant catalytic core of R366C and R366H to bind single stranded nucleic acids, with FP assays demonstrating up to ~10-fold reduced affinity depending on oligonucleotide length (data not shown). Both ssDNA and ssRNA can bind the allosteric sites of SAMHD1[51], and reduced affinity by R366 mutants *in vitro* may indicate an overlap with the catalytic site as well.

2.4 Future Directions

This functional panel is a surface-level characterization at best. Each experiment can be expanded to ask deeper questions. For example, R366 is newly implicated in single-stranded nucleic acid interactions in our manuscript's fluorescence polarization data[91]. Previous co-crystal structures of SAMHD1 only capture ssDNA/RNA overlapping the tetramer interface and A1/A2[51, 95]. Future work should focus on determinants in SAMHD1 interactions with single stranded nucleic acids and the biological scenarios requiring them. Although nucleic acid

binding is not currently linked mechanistically to cancer, it is conceivable that aberrant binding can result in breaking specific intermolecular interactions that contribute to SAMHD1 (and cellular) dysregulation. Since suppression of innate immunity is independent of dNTPase activity in dividing 293T cells, contrary to nondividing cells[34, 84], expanding work to investigate the differential effects of R366 mutants in cycling versus noncycling cells is also warranted.

2.5 Experimental Procedures

2.5.1 Protein expression and purification

N-terminally 6x His-tagged SAMHD1 (residues 113-626; H206R/D207N) was cloned into pET28b (Kan). Protein was expressed in *E. coli* Rosetta (DE3) cells and induced for 16 hrs at 18 ˚C. Resuspended cells were lysed via microfluidizer (15k psi) and purified through gravity flow Ni-NTA affinity resin and S200 PG gel filtration chromatrography. Collected protein was concentrated, aliquoted, flash frozen in 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol, and 0.02% sodium azide, and stored at -80 ˚C until further use.

2.5.2 Crystallization and data collection

Crystals for the HD domain of SAMHD1 (residues 113-626) R366C/H were obtained using catalytically inactive H206R/D207N constructs with the microbatch under oil method. R366C or R366H (4 mg/mL; 50 mM Tris pH 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP) was mixed with 4 mM dGTP, incubated on ice for 15 min, and added 1:1 with crystallization buffer (100 mM succinic acid-phosphate-glycine (SPG) buffer, 30% w/v PEG1500; Qiagen). R366C crystallized at pH 8.2 in ~2 weeks at room temperature. R366H crystallized at pH 9.0 within 1-2 days at

room temperature. Both constructs were cryoprotected with 25% v/v glycerol and flash frozen in liquid nitrogen. Diffraction data were collected on the NECAT beamline 24-ID-E at the Advanced Photon Source (APS), Argonne National Lab. Data statistics are summarized in Table 2-1.

2.5.3 Structure determination and refinement

Diffraction images were processed using HKL2000[96]. Structures were solved via molecular replacement (Phaser[97]) using PDB 4BZB as a search model. Models were refined through iterative rounds of restrained and TLS refinement (REFMAC5[98]) and model building (Coot[99]). The R366H crystal was twinned with twinning fractions of 0.57 and 0.43 for the (H, K, L) and (H, -K, -H-L) twinning operators, respectively. Residues 278-283 were unresolved in both structures. Refinement statistics are summarized in Table 2-1. Coordinate and structure factors were deposited in the PDB under accession codes 7LTT and 7LU5 for R366C and R366H, respectively.

All functional methods are available in Bowen and Temple, et al[91].

2.6 Contributions

Functional and biochemical work was performed by members of the Kim lab as part of a manuscript in press at *JBC*[91]. Presented here is work by co-first author Nicole Bowen (Figs 1B-D, 2, 3B); Caitlin Shepard (Figure 1A); Mirjana Persaud (Figure 3A); and Adrian Oo (Figure 3C). Fidel Arizaga in the Xiong Lab was significant help in optimizing crystal hits for R366C and R366H mutants and providing company during data processing.

3. Towards a full-length human SAMHD1 structure

3.1 Introduction

Currently no structure of human full-length SAMHD1 exists. Indeed, the SAM domain has not been extensively studied as it has been shown to be dispensable for tetramerization, enzymatic catalysis, RNA binding, and retroviral restriction[32, 88]. Despite up to 74% sequence identity, it has been shown that the contribution of the SAM domain to mouse SAMHD1 tetramerization and activity are markedly different from the human homologue[88, 100]. Intriguingly, full-length human SAMHD1 has decreased dNTPase activity relative to the HD domain alone[88]. In contrast, mouse SAMHD1 tetramer formation, dNTPase activity, and HIV-1 restriction ability are severely impaired unless the SAM domain is present. Thus, it is clear even within highly similar evolutionary cousins that the role of the SAM domain is not well understood.

The SAM domain in mouse SAMHD1 acts as a cap on the allosteric nucleotides and helps pin the tetramer together, but the human SAM domain seems to lack those stabilizing contacts and has remained elusive in structure determination. Since a full-length SAMHD1 tetramer is an ideal size (~290 kDa) and symmetry candidate, we then pursued the first cryo-EM structures of human $SAMHD1_{RN}$ in the hopes of capturing dynamic states of the SAM domain. Furthermore, crystal structures of T592E[44] or *in vitro* phosphorylated SAMHD1[24] do not resolve residues past approximately residue 580, suggesting that the modification may loosen previously ordered C-terminal residues. We therefore included a full-length $SAMHD1_{RN}$ T592D phosphomimetic in our analysis.

3.2 Results

3.2.1 SAMHD1$_{RN}$ and T592D assemble into dGTP-activated tetramers

SAMHD1$_{RN}$ (residues 1-626) was analyzed by gel filtration chromatography and negative stain EM to confirm the ability of the human full-length catalytically-inactive protein to tetramerize. In the absence of dGTP, which is the only dNTP capable of satisfying A1, A2, and the catalytic site, SAMHD1 elutes at a volume consistent with the 72 kDa monomer. Incubation on ice with 2 mM dGTP, orders of magnitude above the reported low micromolar EC_{50} for dNTP-mediated allosteric activation[23], stimulated near complete tetramer formation as observed by a ~2.5 mL earlier peak shift and agrees well with the theoretical elution volume of a ~290 kDa tetramer (Figure 3-1a).

A phosphomimetic mutant, SAMHD1$_{RN}$ T592D, was purified and similarly examined. Hereafter, it is referred to as T592D for simplicity. Unlike SAMHD1$_{RN}$, the T592D tetramer fraction was smaller. Tetramer assembly was also limited by kinetics for both constructs at these concentrations, as incubation with dGTP stimulated more complete tetramerization after 4 hours on ice compared to 30 min incubations (data not shown). Negative stain EM micrographs visualize largely intact globular particles amidst a distribution of oligomeric states (Figure 3-2b), which motivated me to include 2 mM dGTP in the tetramer fraction immediately following column passage to ensure tetramer stability. Assembled tetramers have diameters of 11-12 nm.

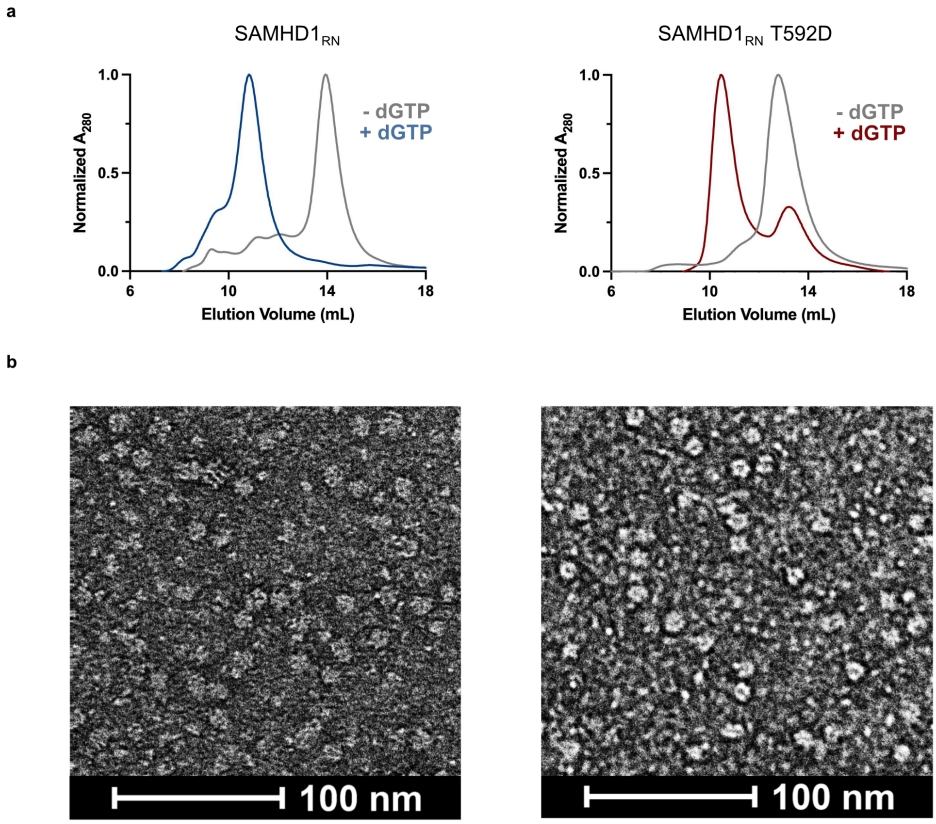

Figure 3-1. Tetramer formation of full-length $SAMHD1_{RN}$ and $SAMHD1_{RN}$ T592D. A) Incubation with 2 mM dGTP stimulates robust tetramerization of $SAMHD1_{RN}$ (left, blue) and the T592D phosphomimetic (red, right). B) Representative negative stain micrographs for each construct (left, $SAMHD1_{RN}$; right, $SAMHD1_{RN}$ T592D).

3.2.2 Preliminary cryo-EM datasets show no observable SAM domain density

Particles were well-behaved and stable with 2 mM dGTP in solution and allowed us freeze and optimize grids. We recently collected initial datasets for both constructs using a 200 kV Glacios cryo-TEM microscope, which was interrupted by a facility power outage for $SAMHD1_{RN}$ analysis. However, a partial dataset for $SAMHD1_{RN}$ was obtained from 928 images, compared to 2484 images for T592D, and allowed us to resolve side chain density in both cases. Figure 3-2a shows a heatmap of the particle orientations for both tetramers in which T592D exhibits more even distribution due to a larger observed number of particles (592,532 as opposed to 176,727). Class averages show no remarkable differences in orientation, yet also show no evidence of an associated SAM domain.

Closer inspection of 2D class averages shows the C-terminal lobe is modestly blurred in some classes for both $SAMHD1_{RN}$ and T592D. Following *ab initio* reconstruction in cryoSPARC, I performed non-uniform refinement with D2 symmetry enforced resulting in a 2.8 Å map for $SAMHD1_{RN}$ and a 2.6 Å map for T592D (Figure 3-2b). Interestingly, the generated maps are practically identical to each other (cross $FSC_{0.143}$ < 3 Å), no SAM domain density is observed, and the maps are thus not significantly different than the X-ray crystallographic structure of dGTP-bound $SAMHD1_{RN}$ HD domain (residues 113-626; Figure 3-2c)[10]. Both full-length $SAMHD1_{RN}$ and its phosphomimetic however exhibit disorder in the C-terminal lobe beyond residue 580, unlike the X-ray model.

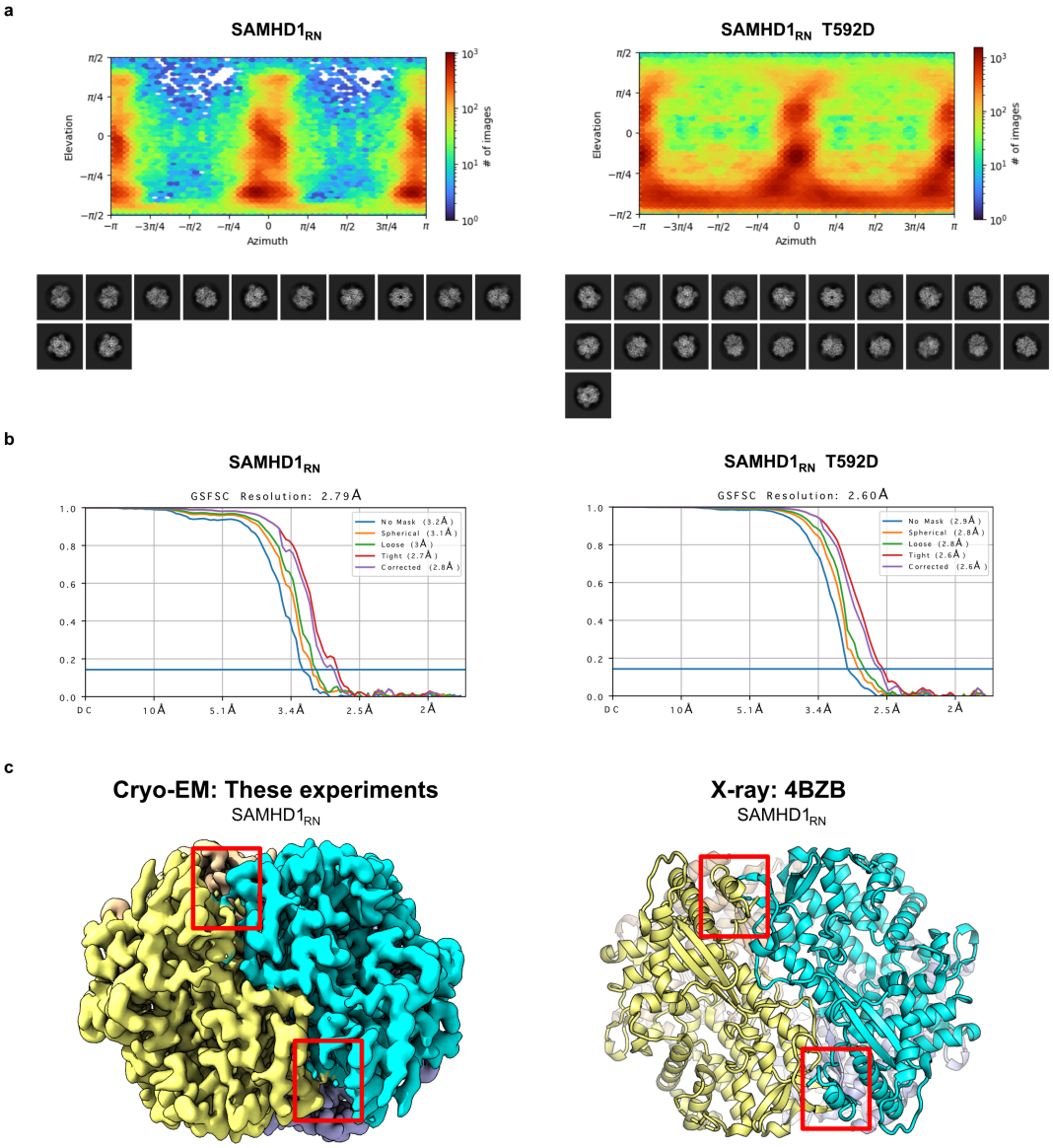

Figure 3-2. Cryo-EM structures of SAMHD1$_{RN}$ and T592D. A) Particle orientation distribution plots and the associated 2D class averages for each construct. B) 3D volume alignment and gold standard Fourier shell correlation (GSFSC) plots showing resolution determination at the 0.143 threshold. C) Comparison between the 2.8 Å cryo-EM map of SAMHD1$_{RN}$ and an X-ray model (PDB code: 4BZB) of dGTP-bound SAMHD1$_{RN}$. Red boxes indicate regions of disorder in both the SAMHD1$_{RN}$ and T592D cryo-EM maps that have been previously resolved for the HD domain alone. No SAM domain density is observed in the cryo-EM maps.

3.3 Discussion

As a globular particle with D2 point group symmetry and labile domains unresolvable by crystallography, we hypothesized that SAMHD1 would be a good candidate for single-particle cryo-EM studies. With this technique we hoped to capture potential dynamic conformational states of the full-length tetramer, but SAM domains remained completely unresolved in our preliminary maps. Notably, the maps for $SAMHD1_{RN}$ and $SAMHD1_{RN}$ T592D were essentially the same. Both structures also showed density for dGTP at the catalytic and A1/A2 sites.

Previous X-ray crystal structures of a T592E phosphomimetic or *in vitro* phosphorylated SAMHD1 lack ordered density extending beyond residue ~580[24, 44], which was also the case here. Figure 3-3a shows a dGTP-bound tetramer rigidly docked into the volume for $SAMHD1_{RN}$. The C-terminal lobe is overall slightly less ordered and no density for the small T592-containing helix or beyond is observed like in previous (bacterially-expressed, non-phosphorylated HD domain) X-ray models[10]. This is true even for $SAMHD1_{RN}$, which lacks modification or mutation at T592. Residues stretching between 116-119 make direct contacts with allosteric nucleotides (Figure 1-2) and density is only resolved starting at D113, akin to previous crystal structures (Figures 3-2c and 3-3a).

The high similarity between $SAMHD1_{RN}$ and T592D maps, combined with all previous structural information, suggests that the phosphorylated protein contains an assembled catalytic core that is largely invariant from the non-modified tetramer. Thus, phosphorylation-dependent regulation of SAMHD1 dNTPase activity, discussed at length in the Section 1.1, may be governed by 1) interactions

with other macromolecules that are dependent on the phosphate moiety of pT592, or 2) the combinatorial influence of pT592 with other modifications on SAMHD1 structure or interactions. Bacterial expression of SAMHD1 served as a useful tool in allowing us to dissect the effects of T592 phosphorylation alone on tetramer structure and confirm results from X-ray crystal structures. SAM domain lability relative to the HD domain also appears unaffected by T592 phosphorylation. These results stand to be confirmed using mammalian-expressed protein, and further work is ongoing regarding isolation of endogenously phosphorylated SAMHD1 as well as SAMHD1 co-immunoprecipitated with binding partners from mammalian cells.

Classifying either data set using up to 3 *ab initio* classes without symmetry enforcement failed to capture a discernible fixed orientation for a SAM domain and 2D averages show no evidence for additional domains beyond the catalytic core. Focused refinement of an additional ~30 Å of space surrounding the current particle boundary (~110 Å in diameter) would be ideal to visualize the SAM domain independent of its orientation to the HD domain. However, this is not possible due to the very small (~12 kDa) SAM domain which generates indistinguishable signal in our (and likely any) current microscope and detector setup. Since a single *ab initio* class gives the best map and no evidence for SAM density is observed at low map contours for $SAMHD1_{RN}$ and T592D, the SAM domain is likely highly flexible and free from stable intra- or inter-protomer association within each tetramer.

It is interesting to speculate on the origins of flexibility. Alignment of mouse (mSAMHD1) and human (hSAMHD1) N-terminal sequences are shown in Figure

3-3b. The mouse orthologue is extremely conserved relative to the human version with most variability arising due to the SAM domain. Pairwise sequence identity for the HD domains is 72% (using the EMBOSS alignment server[101]), whereas the SAM domains reside at 50% (and drop to ~37% if including the 31-residue N-terminal addition in mouse). One region contributing to this variability involves residues 100-112 in hSAMHD1 (corresponding to mouse positions 132-144) immediately preceding D113, which warrants investigation as to its role in SAM domain flexibility.

C-terminal flexibility is also apparent (Figure 3-3b). Notably, a helix containing T592 that has been observed in X-ray data of HD domains is absent in our structures. This may reflect the heterogeneity in C-terminal lobe conformations induced by a lack of crystal packing, loosened tetramer packing specific to $SAMHD1_{1-626}$ instead of $SAMHD1_{113-626}$, or a combination of these factors. T592 is also among ~40 C-terminal residues for which no density has been resolved, which delineates an unstructured C-terminal tract capable of extending across the diameter of the tetramer core. Numerous critical modifications occur in this tail, including phosphorylation, SUMOylation, and ubiquitination (Figure 1-1) which require direct interaction with the appropriate enzymes and have major functional consequences. No gross structural deviations caused by T592D were revealed here, but pT592-based regulation of catalytic activity may still be dependent on a disordered-to-structured transition (or vice-versa) in unresolved regions of these maps induced by the modification, as observed for other phosphorylation-regulated proteins[102].

Thus, bacterially-produced full-length $SAMHD1_{RN}$ and $SAMHD1_{RN}$ T592D retain similar structure of the catalytic core. In both cases, the SAM domain is a dynamic appendage and the unstructured C-terminal domain, which is tagged in multiple locations, is capable of sampling a large radius for potential interactions. These new cryo-EM data confirm that phosphorylation-dependent regulation of full-length SAMHD1 does not involve pT592-induced structural changes in the catalytic core, or stable intra-tetramer SAM domain interactions. Instead, the N-terminal SAM domain and the C-terminus may engage in PTM-modulated macromolecular interactions that govern SAMHD1 enzymatic regulation. The C-terminal lobe lies directly adjacent to a cluster of regulatory cysteines (Figure 2-5), sits proximal to space a neighboring SAM domain can sample, and exhibits a 40-residue disordered stretch at the C-terminus that is capable of being multiply-modified and spanning the tetramer's diameter. Considering the four-fold redundancy of these potential interactions in an assembled tetramer, and how they might affect allosteric activation, the opportunity for regulatory crosstalk involving these regions is evident.

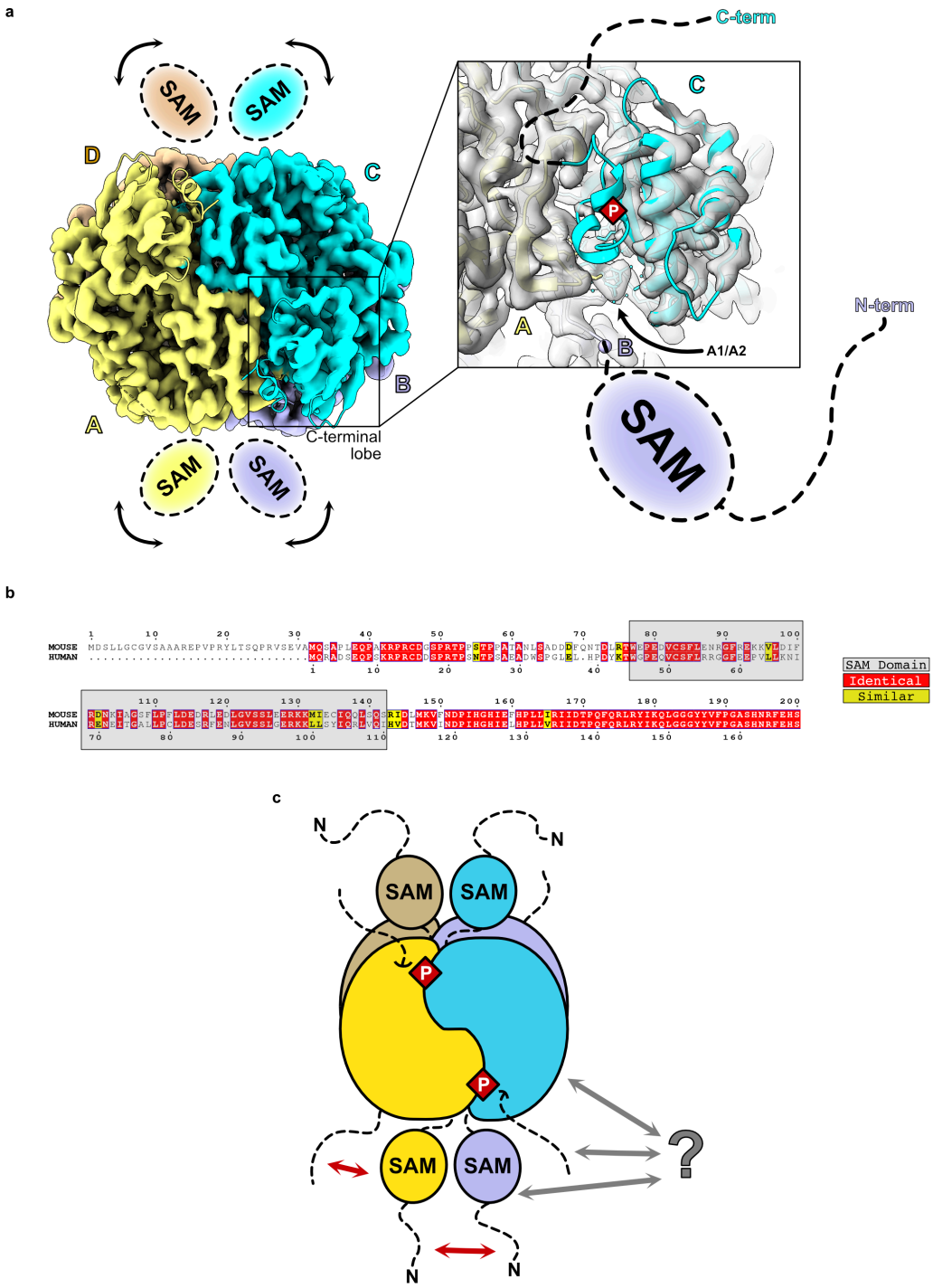

Figure 3-3. N-terminal SAM domains and C-termini are highly flexible. A) PDB: 4BZB rigid docked into the SAMHD1$_{RN}$ map with each subunit colored and labeled accordingly. The inset shows the disordered C-terminal lobe sitting over the A1/A2 allosteric nucleotide binding sites. Red diamond indicates the location of T592 phosphorylation. The sphere indicates the first ordered N-terminal residue of SAMHD1 (D113) observable by both X-ray crystallography or cryo-EM, which is situated directly adjacent to A1/A2. B) Sequence alignment (generated using the

ESPript server[103]) of the mouse and human SAMHD1 N-terminus. The SAM domain is boxed in gray. Identical residues are colored red. Similar residues are colored yellow. Residue numbering is indicated above (mouse isoform 1) or below (human isoform 1) the sequence. C) Schematic of a full-length SAMHD1 tetramer. "N" signifies the N-terminus. Relative positions of T592 phosphorylation and the SAM domains are shown. Red arrows indicate the opportunity for intra-tetramer interactions between unstructured N- and C-termini. Gray arrows indicate the possibility for undiscovered interactions with the N-terminal SAM domain, the highly modified C-terminus, and/or the HD domain.

3.4 Future directions

The lateness with which my work surrounding SAMHD1 began, in combination with limitations due to the COVID-19 pandemic, caused the desired progress on this project to be delayed. Currently we are moving forward with multiple prongs of work on SAMHD1 structure. We are interested in full-length WT SAMHD1 structure as opposed to the HD/RN mutant in these studies. We have preliminarily explored general chemical crosslinkers such as glutaraldehyde to limit protein dynamics and attempt to capture regions of conformational space the SAM domain can sample. These studies are ongoing and will include side-chain specific crosslinker screening. A more attractive direction is obtaining SAMHD1 in complex with a binding partner that may stably interact with and lock down the flexible SAM domain. Furthermore, these initial studies were performed with bacterially-expressed SAMHD1, but must be repeated with protein produced in mammalian cells before any conclusions can be drawn for human SAMHD1 structural mechanisms. We are currently collaborating with Li Wu's lab at the University of Iowa to isolate pSAMHD1 from 293T cells with co-immunoprecipitated NF-κB, and I will expand this work to purify (non-phosphorylated) SAMHD1 from non-cycling cells. Further optimization of tetramer stability using different dNTP combinations,

and quantitatively assessing full-length protein assembly kinetics, may also provide improved particles and resolution. Ultimately, data collection on a 300 kV Titan Krios is planned.

3.5 Experimental procedures

3.5.1 Protein expression and purification

N-terminally 6x His-tagged SAMHD1 (residues 1-626; H206R/D207N +/- T592D) was cloned into pET28b. Proteins were expressed in *E. coli* Rosetta (DE3) cells and induced at OD ~ 0.6-0.8 for 16 hrs @ 18 ˚C. Resuspended cells (Ni_A buffer: 50 mM Tris pH 8.0, 500 mM NaCl, 0.1 mM TCEP, 5% v/v glycerol, and 10 mM imidazole) were lysed via microfluidizer (15k psi) and purified through gravity flow Ni-NTA affinity resin and S200 PG 16/60 gel filtration chromatography. Collected protein was concentrated, aliquoted, flash frozen in 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol, and 0.02% sodium azide, and stored at -80 ˚C until further use.

3.5.2 Tetramer assembly

Tetramers were assembled fresh from thawed protein. $SAMHD1_{RN}$ (6.6 mg/mL) or $SAMHD1_{RN}$ T592D (12.4 mg/mL) was combined with dGTP (2 mM) and gel filtration buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP) and incubated on ice for at least 4 hrs, agitating every hour. Samples were applied to a S200 Increase GL 10/300 gel filtration column and tetramer fractions were immediately spiked with 2 mM dGTP after coming off the column to preserve tetramers until grid preparation.

3.5.3 Negative stain EM grid preparation and data collection

Generally, protein (~0.05 mg/mL; 4 µL) was applied to grids (Cu 400 mesh with carbon film; Electron Microscopy Sciences) pretreated by glow discharge for 15 s at 10 mA (Gloqube Plus; Electron Microscopy Sciences). Following 1 min incubation, liquid was blotted away and sample was briefly massaged in 15 µL of 2% uranyl acetate, (UA) blotted, then dipped in another UA drop and incubated for 1 min. All stain was blotted and the grid was allowed to air dry. Images were collected using a Talos L120C equipped with a BM-CETA CCD detector.

3.5.4 Cryo-EM grid preparation and data collection

Tetramer fractions (~350 mAU) stabilized with dGTP were applied (3 µL) to a 300 mesh C-flat 2/1 Cu grid (Electron Microscopy Sciences) pretreated by glow discharge for 15 s at 10 mA. Each grid was blotted at 20 ˚C, 100% humidity for 1 s with blot force of -1, plunge-frozen into liquid ethane using a FEI Vitrobot Mark IV (Thermo Scientific), and stored in liquid nitrogen until data collection.

Images were collected using a Thermo Scientific Glacios electron microscope equipped with a 200 kV FEG and a Gatan K2 direct electron detector. Raw pixel size was 0.4515 Å. A total of 928 images (cut short due to Hurricane Ida's power outage!) for $SAMHD1_{RN}$ and 2484 images for $SAMHD1_{RN}$ T592D were collected. A defocus range of -1 to -2 µm was used and data were collected with a total dose of 41 e/Å^2, or 4.65 s over 31 frames. Standard data processing procedures in cryoSPARC were used[104], including non-uniform refinement with D2 symmetry enforced. The final global resolution was 2.8 Å from 176,727 particles for $SAMHD1_{RN}$ and 2.6 Å from 592,532 particles for $SAMHD1_{RN}$ T592D. The

crystal structure of dGTP-bound SAMHD1$_{RN}$ from PDB: 4BZB was rigid docked into the maps using ChimeraX without need for further fit refinement.

3.6 Contributions

Many thanks to Kaifeng Zhou, Wei Zheng, and Jianfeng Lin who taught me, helped prepare grids, and collected data.

4. Appendix I: Probing HIV-1 capsid stability and interactions

4.1 Introduction

As a retrovirus, HIV-1 is tasked with reverse transcribing its ssRNA genome into double-stranded complementary DNA (cDNA), translocating it to the nucleus, and integrating into the host genome. The viral genome is contained within a proteinaceous capsid – a protective cargo vessel porous to nucleotides[105], doubling as a reaction chamber for reverse transcription – while it is trafficked to the nucleus along microtubule networks. The capsid is an important biological structure that must protect the vulnerable viral genome from restriction factors, interact with necessary trafficking machinery including the nuclear pore complex (NPC), and disassemble at the proper moment to allow for genome integration, as we review in[106].

Each capsid is composed of many copies of the capsid protein (CA) which oligomerizes into hexamers (~250) and pentamers (12) that form the conical shell[107, 108] (Figure 4-1a). Numerous proteins must interact with the assembled capsid to facilitate productive infection and the capsid has thereby evolved to recognize various common motifs, which can be studied through the assembly of $CA_{A14C/E45C}$ tubes that recapitulate the pattern of the capsid surface (Figure 4-1b). Patterns formed at four main interfaces are propagated throughout the lattice, each with unique charge distributions, and serve as signposts for capsid-binders (Figure 4-2). My work regarding CA focused on interactions at the central pore, which is most notably characterized by a 6-fold ring of R18 residues (Figure 4-2). This electropositive channel makes the capsid porous and can coordinate phosphate

moieties of important small molecule metabolites including dNTPs to fuel reverse transcription and IP6 which seems important for DNA synthesis and stability[105, 109, 110]. Other polyanions located on disordered protein domains are capable of interacting with the central pore, including fasciculation and elongation protein zeta-1 (FEZ-1) which acts as a kinesin-1 adaptor and participates in capsid translocation to the nucleus[111]. Tracts of sequential polyglutamate (pE) residues have been shown biochemically and computationally to mediate binding to the central R18 ring in CA hexamers[112]. Other proteins bearing pE motifs or modifications then became my focus, including polyglutamine binding protein 1 (PQBP1, a nuclear transcription/splicing factor that binds polyglutamine repeats[113]) and cyclic GAMP synthase (cGAS) discussed below.

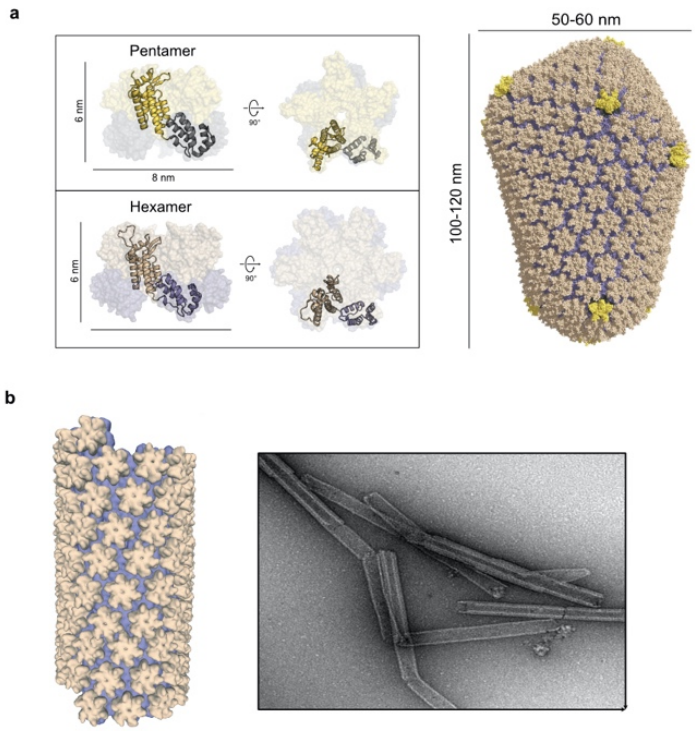

Figure 4-1. Organization of HIV-1 CA assemblies. A) The HIV-1 capsid is composed of 12 pentamers (yellow and gray, top) and over 200 hexamers (wheat and slate, bottom). Model is based off of PDB 3J3Q. B) Model of helical disulfide-stabilized $CA_{A14C/E45C}$ tubes and visualization through negative stain EM micrograph.

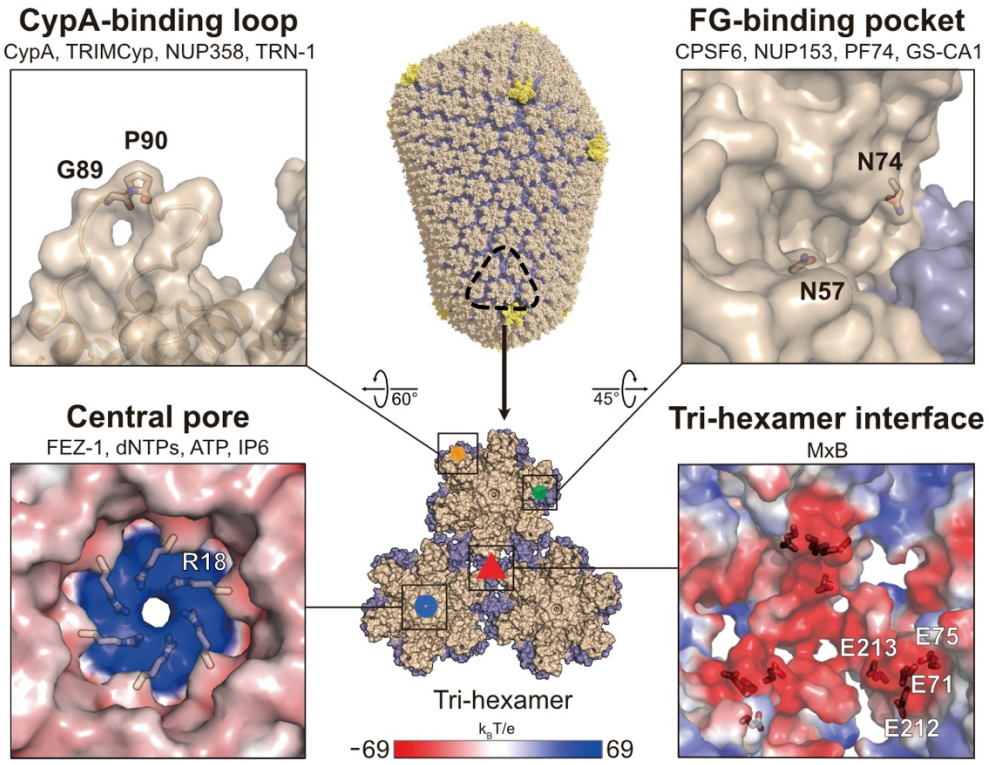

Figure 4-2. Commonly targeted sites of the HIV-1 capsid surface with important residues noted and relevant binding factors listed. Figure adapted from[106].

Initial work began with cGAS, a cytosolic DNA sensor that mediates innate immune responses by catalyzing formation of cGAMP (cyclic 2'-3'-guanosine-adenosine monophosphate) to mount IFN-I responses[59, 114] and positively regulate T-cell activation[115] upon activation with dsDNA[116] (Figure 4-3). Contact with cytosolic DNA causes liquid phase separation in which cGAS crosslinks nonspecifically with DNA and condenses into highly enzymatically active populations, which concentrates cGAMP product formation to activate downstream innate immune activation[117]. Due to low observed DNA affinity and relative binding promiscuity, co-receptors have been proposed as a mechanism of specificity and affinity; PQBP1 is a co-receptor for cGAS and directly binds the HIV-1 RT product to help induce an antiretroviral immune response[118]. PQBP1

possesses a disordered N-terminal poly-acidic patch, and cGAS (as well as tubulin[119]) is post-translationally polyglutamylated[120]. Although little is understood about the general nature of polyglutamylation as a PTM, it can affect processivity and velocity of tubulin-traversing microtubule motors[121] and negatively regulate the activation state of cGAS[120]. Building off of our lab's previous work with FEZ-1, I show a preliminary identification of polyglutamate as a motif directing capsid-binders to the central hexamer pore.

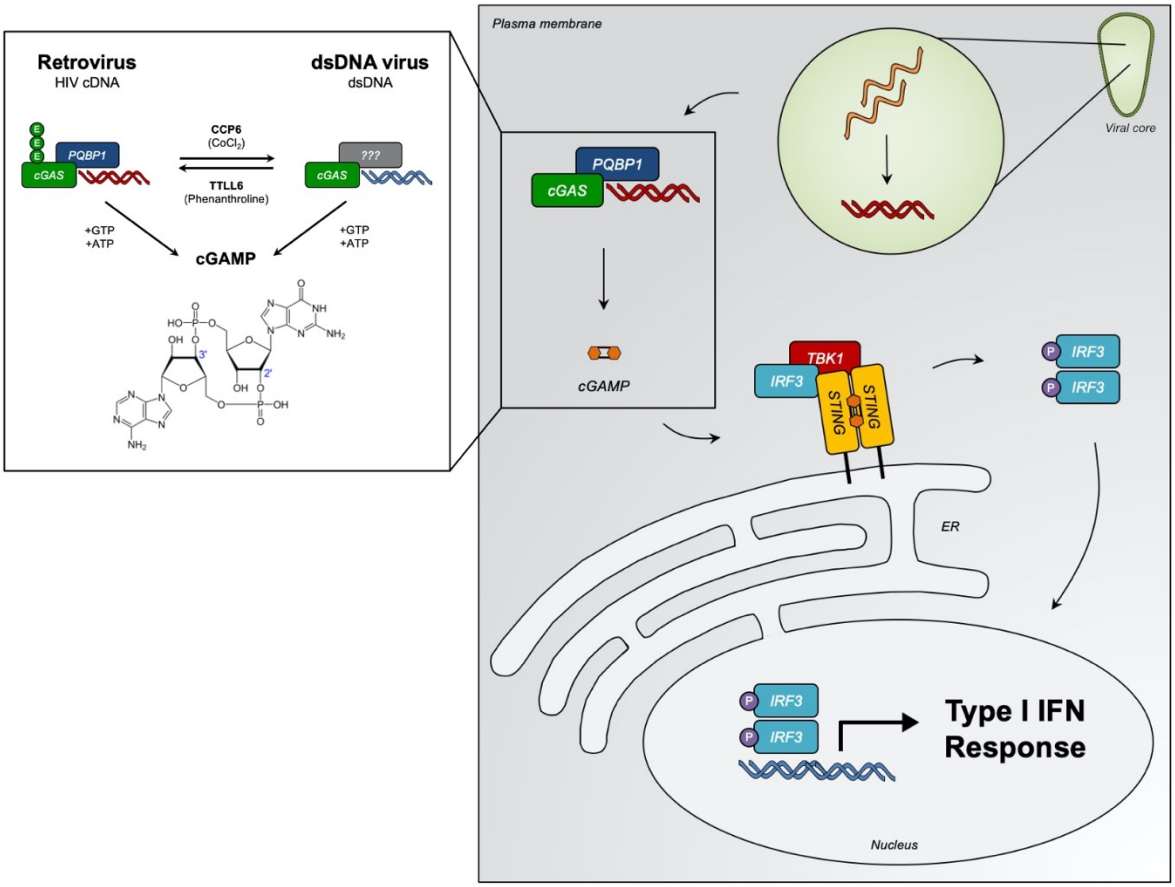

Figure 4-3. Illustration of cGAS-mediated immunity in which, following cellular deposition, retroviral RNA is reverse transcribed into cDNA and stimulates cGAS/PQBP1-mediated activation of innate immunity through cGAMP synthesis. A working model based on collaborator data was that PQBP1 can only co-stimulate innate activation if cGAS was pE-modified (flyout).

4.2 Results

4.2.1 Purification and binding of cGAS and PQBP1

cGAS has been reported to form soluble aggregates with DNA[117]. I recombinantly overexpressed full-length human cGAS (residues 1-522) and $cGAS_{\Delta NTD}$ (residues 161-522) with pRARE and +/- TTLL6, the endogenous polyglutamylase of cGAS[120]. Large soluble aggregate species are observed regardless of sequence length or pE presence, but pE modification through TTLL6 coexpression markedly reduced the $A_{260/280}$ of the DNA-associated aggregates from 1.46 to 1.11 (and increased the proportion of smaller oligomers ($A_{260/280} \approx 0.8$-0.9; Figure 4-4a). Later cGAS preps including DNase treatment achieved monodisperse monomeric protein that was prone to degradation in solution. PQBP1, which is mostly disordered, was purified and runs as a dimer according to size on gel filtration. No pairwise affinity between $cGAS_{+/-NTD}$ and PQBP1 is observed at low μM concentrations on gel filtration (Figure 4-4b).

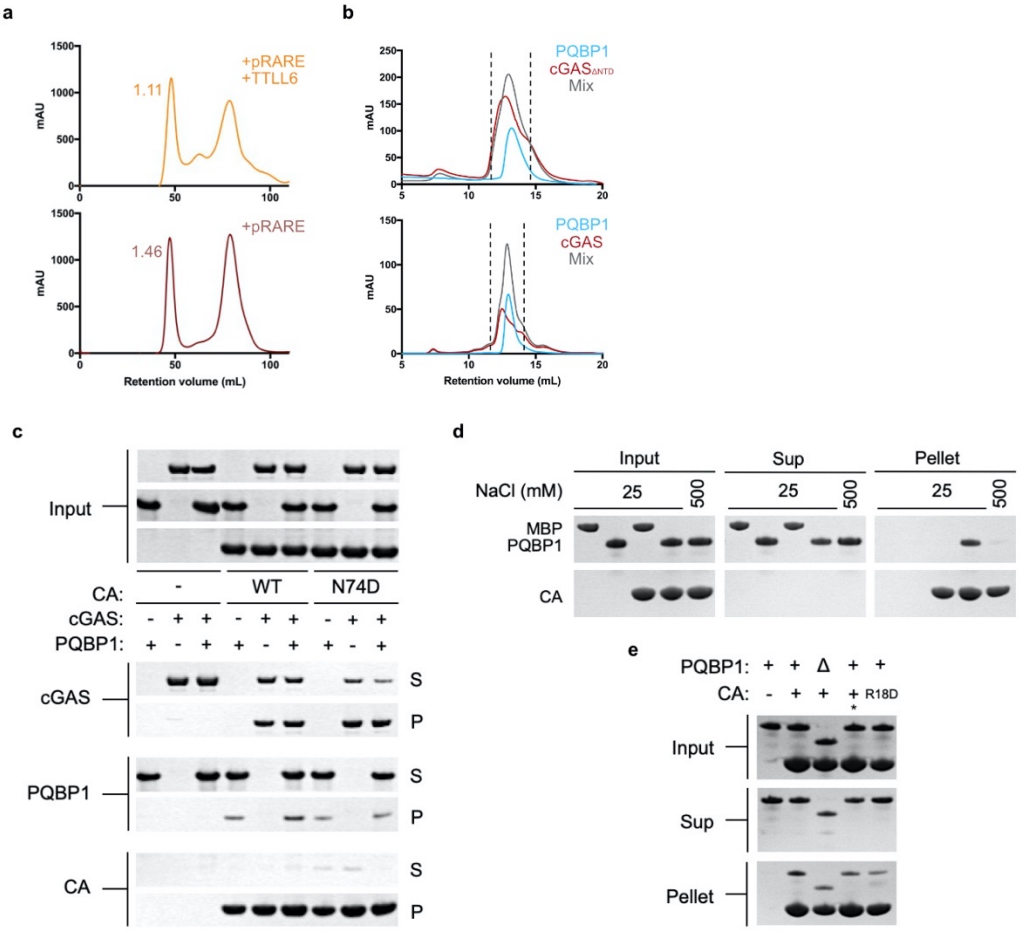

Figure 4-4. Expression and binding of cGAS and PQBP1 constructs and interaction with CA. A) Co-expression of cGAS with TTLL6 reduces the $A_{260/280}$ of the void peak, presumably soluble nonspecific DNA-cGAS aggregates. B) Initial binding (shown) and subsequent "cleaner" binding with monodisperse cGAS (not shown) show no significant interaction by gel filtration chromatography, but this bears repeating. C) PQBP1 and cGAS have non-additive affinities for CA tubes, with largely similar affinities between WT and N74D tubes. D) PQBP1 pelleting with CA tubes is salt-sensitive. MBP: maltose binding protein (negative control). E) PQBP1 pelleting is sensitive to the present of the NTD and charge neutralization at the R18 hexamer pore.

4.2.2 Interaction of cGAS and PQBP1 with CA constructs

Neither cGAS nor PQBP1 interacted with soluble A14C/E45C/W184A/M185A crosslinked CA hexamers (data not shown). However, both proteins demonstrated affinity for helical arrays of HIV-1 CA tubes in a pelleting assay that takes advantage of the insolubility of assembled A14C/E45C

crosslinked CA. Neither PQBP1 nor cGAS showed increased or diminished pelleting with tubes in a mixed sample (Figure 4-4c). Focus of our collaboration shifted to PQBP1, which seemed to be recruited first to decorate the HIV-1 capsid upon cellular deposition. I further characterized PQBP1 binding to tubes. PQBP1 has concentrated acidic residues, particularly rich in glutamates, on its NTD (residues 1-46) and overall had weak affinity for CA tubes at 75 mM NaCl. Stronger binding is observed at 25 mM NaCl (4-4d).

I hypothesized that the higher negative charge density of the NTD (Figure 4-5a) targets the central pore of assembled CA hexamers. Consistent with this hypothesis, 500 mM salt abrogated the expected electrostatic interaction between the electropositive R18 pore and acidic PQBP1 residues and slight reduction in binding can be observed using $PQBP1_{\Delta 1\text{-}46}$, IP6 addition, or CA_{R18D} tubes (Figure 4-4e). Alanine-scanning mutants were used to further identify regions of binding significance, but pelleting assays were generally not sensitive enough to capture subtle differences in weak binding. Protein ended up being sent instead for spectroscopic binding assays in collaboration with the Chanda group at SBP and Böcking group at UNSW, in press at the time of this writing[122]. Indeed, these assays show the N-terminal 46 residues of PQBP1 are sufficient for strong capsid binding, and that PQBP1 binding is abrogated by hexacarboxybenzene addition (HCB, a polyanionic R18 pore ligand) and CA_{R18G} mutation (manuscript in press).

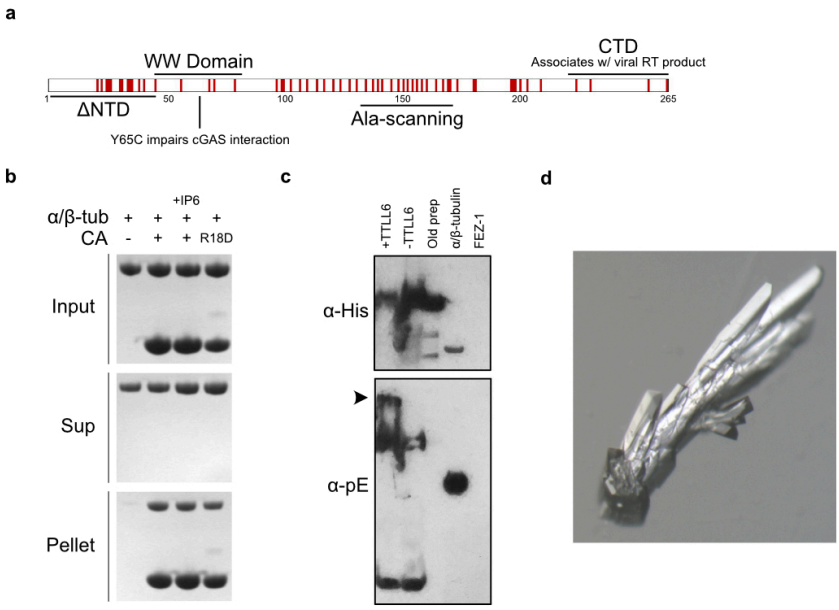

Figure 4-5. pE motifs are common capsid binders. A) Schematic of PQBP1 primary structure with all acidic residues highlighted in red. Structured domains are noted on top and relevant constructs to my experiments are denoted below the schematic. B) Tubulin pelleting assay with disulfide-stabilized CA tubes. C) Western blot to detect pE modification on various preps of full-length cGAS, with tubulin as a positive control and FEZ-1 as a negative control. Arrow indicates unique band in TTLL6-coexpressed protein. D) Crystal obtained from 6x pE peptide additive screen with disulfide-stabilized HIV hexamers. No density was observed for the peptide.

4.2.3 Crystallization of CA hexamers with a pE peptide is unfruitful

We have known HIV-1 capsid cores are trafficked along tubulin-based microtubule networks, yet pelleting assays I performed positively demonstrated direct interaction between natively purified calf brain α/β-tubulin and CA tubes (Figure 4-5b), which is one of few other relatively well-studied pE-modified proteins[119]. IP6 and CA_{R18D} tubes however seemed to have no effect on tubulin pelleting, which remained strong, indicating further important binding surfaces. However, it is clear that other proteins that bear polyglutamate motifs such as FEZ-1, PQBP1, and cGAS which is post-translationally modified by TTLL6 (which is confirmed via Western blot in Figure 4-5c) share a common targeting mode. I

attempted to crystallize an HIV-1 $CA_{A14C/E45C}$ hexamer with a 6x-polyglutamate peptide. Extremely large crystals formed that were stable for weeks (Figure 4-5d). A diffraction data set was collected revealing a novel (P1) space group, with four hexamers in the asymmetric unit but ultimately no additional density for a pE peptide (data not shown).

4.3 Discussion

Coexpression of TTLL6 with cGAS, thereby increasing the pE-modified population, seems to decrease the amount of nucleic acid bound in void species that elutes during cGAS size exclusion purification (Figure 4-4a). Unpublished collaborator data suggested cGAS may need to be pE-modified to act with PQBP1 to mount an immune response (Figure 4-3). No pairwise interaction could be observed between PQBP1 and cGAS with or without pE modification (Figure 4-4b; data not shown) but interaction has been shown by co-IP previously[118]. Although neither cGAS nor PQBP1 could interact with soluble CA hexamers at low μM concentrations by gel filtration, both showed affinity at similar concentrations for CA tubes in pelleting assays. PQBP1 affinity can be mostly attributed to its acidic, negatively charged N-terminal 46 residues acting on the central hexamer pore, but less important binding surfaces may still be involved. This is supported by PQBP1 showing modest yet repeatable decreased pelleting with N74D tubes, a hyper-stable capsid mutation which might implicate the FG-binding pocket (Figure 4-2). Binding between cGAS and PQBP1 appeared independent in part due to their inability to saturate the numerous binding sites available on CA tubes at low micromolar concetrations. Although crystallization with a pE peptide showed CA

hexamers in a P1 space group for the first time in our lab, no density was observed for the peptide at the central pore or elsewhere.

4.4 Future Directions

Further work must be completed to ascertain if PQBP1 and cGAS independently bind the CA lattice or participate in a ternary complex, which was not observed here. Interestingly, PQBP1 binds the HIV-1 RT product, Determinants in PQBP1 binding are now well-understood, structural analysis is precluded since PQBP1 is largely disordered[123]. Further work should be devoted to uncovering determinants in cGAS association with CA tubes. As both cGAS and PQBP1 can associate with the viral cDNA[118], the potential role of nucleic acid in complex with these proteins and the capsid should be evaluated.

4.5 Experimental Procedures

4.5.1 Protein expression and purification

Constructs for full-length (1-522) and ΔNTD (161-522) cGAS were cloned into pET16 vectors (Amp) with an N-terminal 6xHis-MBP tag and intervening TEV protease cleavage site. Sequences encoding residues 1-265 and 47-265 of PQBP1 were inserted into MCS1 of pCDFDuet-1 with a 6x N-terminal His-tag. Similarly, full-length human TTLL6 was inserted into MCS1 of pRSFDuet-1 with an N-terminal 6x His-tag. Proteins were expressed in *E. coli* BL21 (DE3) cells (or the Rosetta cell line and co-transfected with pRARE for cGAS expression) and induced at OD_{600} ~ 0.8 for 16 hrs at 18 ˚C. Cells were resuspended and lysed in Ni_A buffer supplemented with a protease inhibitor cocktail tablet (Roche) using a microfluidizer (15k psi). 6xHis-PQBP1 was purified through Ni-NTA affinity and

S200 PG 16/60 gel filtration chromatrography. Both cGAS constructs were treated with DNAse 30 min prior to lysing, then purified through Ni-NTA affinity and S200 PG 16/60 gel filtration, with an optional HiTrap QP anion exchange step.

$CA_{A14C/E45C}$-[mPro]-MBP-6xHis and $CA_{A14C/E45C/W184A/M185A}$ were cloned into pET11a (Amp). Protein was expressed in *E. coli* BL21 (DE3) cells and induced at OD_{600} ~0.6-0.8 at 25 ˚C for 16 hrs. Cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.1 mM TCEP, 5% v/v glycerol), lysed via microfluidizer (15k psi), and clarified at 15000 rpm for 35 min. Samples were purified separately. $CA_{A14C/E45C}$-[mPro]-MBP-6xHis was purified through gravity Ni-NTA affinity chromatography followed by concentration and dilution into Q_A buffer (50 mM Tris pH 8.0, 0.1 mM TCEP) and passage through a HiTrap QP ion exchange column. $CA_{A14C/E45C/W184A/M185A}$ supernatant was subject to a 25% ammonium sulfate precipitation at 4 ˚C for 2 hrs while stirring but minimizing bubbles. The precipitation was centrifuged at 15000 rpm for 20 min and the pellet was resuspended slowly and dialyzed overnight in excess S_A buffer (25 mM HEPES pH 7.5, 0.1 mM TCEP). Protein was further purified through HiTrap SP ion exchange chromatrography. All protein was concentrated in 50 mM Tris pH 8.0, 75 mM NaCl, and 40 mM βME and flash frozen for storage at -80 ˚C until use.

4.5.2 Western blots

Following SDS-PAGE, protein bands were transferred to a PVDF membrane via electrophoresis, blocked with milk, and incubated with the appropriate antibody according to standard protocol. Either mAb GT335 (Adipogen; detects the branch points on glutamate side chains of all forms of

glutamylated protein), pAb IN105 (Adipogen) which detects polyglutamate chains, were used as primary antibodies)[124], and α-His (from where??) were used as primary antibodies. Orthogonal secondary antibodies conjugated to horseradish peroxidase (HRP) were used with ECL reagent (Thermo Fisher) according to manufacturer's protocol and exposed to film for detection.

4.5.3 Gel filtration binding assays

Binding reactions were typically performed at 75 mM NaCl with 50-100 μM binding partners allowed to incubate on ice for 30 min to 1 hr. Samples were applied to the appropriate gel filtration column, typically a Superdex resin like S200 GL Increase 10/300, in a running buffer made of 50 mM Tris pH 8.0, 75 mM NaCl, and 0.1 mM TCEP, and elution profiles were analyzed.

4.5.4 Pelleting assays

Disulfide-stabilized CA tubes were formed by dialyzing $CA_{A14C/E45C}$ (~15 mg/mL) at 4 ˚C overnight in 50 mM Tris pH 8.0, 1 M NaCl, and 40 mM β-mercaptoethanol, followed by two nights in 50 mM Tris pH 8.0, 1 M NaCl, and finally overnight in 50 mM Tris pH 8.0. Concentration of CA tubes was determined using A_{280} and monomeric $Abs_{0.1\%}$ = 1.31. Pelleting reactions were performed at 25 mM NaCl and mixing 100 μM CA (16.7 μM hexamers) with typically 50 μM binding factor unless otherwise noted. Reactions were incubated at room temperature for 30 min and pelleted by tabletop centrifugation at 4 ˚C at 14000 rpm for 15 min. Supernatants were carefully removed from the pellet and their volume estimated to resuspend the pellet. Gel samples for input, supernatant, and pellet fractions were taken at appropriate times and analyzed via SDS-PAGE.

4.5.5 Crystallization

$CA_{A14C/E45C/W184A/M185A}$ (300-400 μM) was assembled into hexamers by dialyzing overnight at 4 ˚C in 50 mM Tris pH 8.0, 1 M NaCl, and 40 mM β-mercaptoethanol, followed by two nights in 50 mM Tris pH 8.0, 1 M NaCl, and finally overnight in 50 mM Tris pH 8.0. Concentrations of CA hexamers was determined using A_{280} and monomeric $Abs_{0.1\%}$ = 1.31. Co-crystals for disulfide-crosslinked hexamers with a 6x-pE peptide (Genscript) were obtained using the microbatch under oil method. A solution of hexamers (20 μM; in 50 mM Tris pH 8.0, 75 mM NaCl) was mixed with a 10X molar excess EEEEEE peptide (200μM; resuspended according to manufacturer protocol in 50 mM Tris pH 8.0, 60 mM NaOH) and incubated at room temperature for 30 min before setting trays with 1 μL of protein solution with 1 μL of condition in the crystallization drop. Crystals appeared within 3 days, with the most significant hit in 0.1 M MMT buffer, 25% v/v PEG 1500 (NeXtal PACT D3). The crystal was broken apart and cryoprotected with 25% v/v glycerol and flash frozen in liquid nitrogen. Diffraction data were collected on the NECAT beamline 24-ID-E at the Advanced Photon Source (APS), Argonne National Lab. No structural data are presented since no pE density was observed.

4.6 Contributions

Some of the work shown here references data recently submitted for publication[122] in collaboration with Sumit Chanda's group at SBP Medical Discovery Institute.

5. Appendix II: Prototyping DNA origami scaffolds to mitigate the air-water interface in cryo-EM grid preparation

5.1 Introduction

As the sophistication of cryo-EM processing capabilities have developed, particle adsorption to the air-water interface (AWI) has had little match in the bottleneck it presents for routine high-resolution structure determination. AWI-mediated denaturation of proteins is to an extent unavoidable. Following blotting of a sample from a grid, the blotted grid is exposed to air while quickly plunged into liquid ethane to form vitreous ice, which is necessary for electron-transparency in contrast to crystalline ice[125, 126]. Residual protein deposited in holes of the grid has a high surface area:volume ratio and generally capable of diffusing to the AWI in under 1 ms[127]. Hydrophobic cores of folded proteins tend to associate with the hydrophobic AWI, which leads to partial denaturation and/or anisotropic preferred orientations of particles relative to the AWI[125] and can affect upwards of 90% of particles[128]. Thus, the AWI can introduce major deleterious effects during atomic resolution cryo-EM structure determination[129, 130].

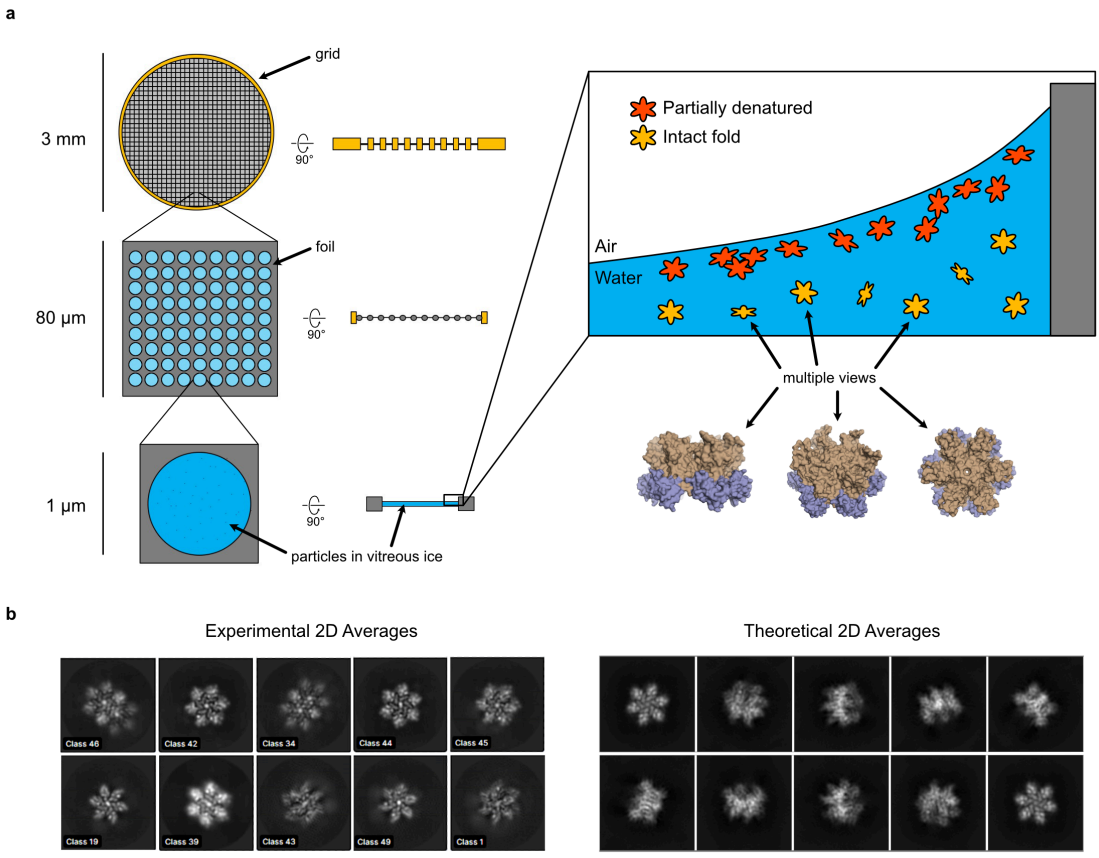

Figure 5-1. Grid structure and the CA hexamer orientation bias. A) Basic dimensional cryo-EM grid organization (left) and illustration of the AWI problem in single particle analysis (right). Hydrophobic cores of proteins associate with the air interface, which is hydrophobic compared to aqueous solution. This results in partial denaturation of the majority of deposited protein, limiting overall maximum resolution and contrast, and is often coincident with an induced particle orientation bias (red particles, fewer views amongst more aggregation and/or denaturation). B) Experimental 2D class averages of an HIV CA hexamer showing orientation bias versus theoretical class averages showing multiple views.

Various groups have attempted to mitigate the AWI problem at each step of sample preparation, reviewed well by Weissenberger et al[131]. Alternate blotting parameters and faster plunging are obvious avenues to reduce diffusion time and therefore denaturation at the AWI[126]. Additives and detergents have successfully been used in some cases to line the AWI and prevent particle adsorption,

specifically with nonionic and zwitterionic species[132-134], but detergent selection is case-by-case and may still affect native protein structures. Further aspects of sample buffer and construct design can be optimized[131]. Support structures that sequester the protein within the aqueous volume – away from the AWI – are more popular. For example, carbon-based thin films of graphene-oxide[135] or graphene monolayers[136, 137] hydrophilized with plasma have been used to achieve near atomic resolution of a 52 kDa streptavidin particle in the latter case. However, these supports can introduce background signal that makes particle visualization more difficult. Some carbon films can also be chemically functionalized to retain particles away from the AWI, for example using ssRNA to enrich low abundance biological complexes on the grid[138]. Other reports detail using an additional highly stable gold grid, across which a graphene monolayer is supported, to achieve improved image quality and resolution[139].

Another more tunable approach uses DNA origami to serve as a support structure. DNA origami, as its name suggests, refers to folding DNA duplexes into large nanoscale structures and relies on a long ssDNA template strand and numerous oligonucleotide “staples” that govern its three-dimensional organization[140]. Martin et al used a hollow hexagonal origami tube with an internal anchoring point as a scaffold for the transcription factor p53 which recognized specific dsDNA sequences engineered into the scaffold[141]. This scaffold design did not generate multiple orientations as intended, which are needed for high resolution reconstruction and was further limited in application to DNA-binding proteins. In collaboration with Chenxiang Lin’s lab, we worked to develop a

modular DNA origami scaffold to sequester the protein of interest (POI) away from the AWI and allow for multiple orientations to be sampled in order to aid in structural analysis of challenging POIs that exhibit AWI association or preferred orientation issues.

Anecdotally in the Xiong lab, the CA hexamer (described in Appendix I) is notorious for preferred orientation bias while preparing cryo-EM grids (Figure 5-1b). It therefore served as a useful test case in assaying whether our origami construct: 1) easily and efficiently tethered a POI, 2) sequestered the POI away from the damaging AWI, and 3) allowed the particle to freely tumble in solution and sample multiple orientations. Here I summarize results of an extensive negative stain EM screening effort after scaffold prototyping as the first steps in methods development aimed at mitigating effects of the AWI.

5.2 Results

5.2.1 Our origami scaffold prototype is modular and polymerizeable

In collaboration with the Lin lab, we screened methods of attaching a CA hexamer to multiple scaffold designs. Initial work focused on the modularity of the origami monomer. Our design included the capacity for axial polymerization, which we hypothesized would allow us to control monomer stacking and therefore regulate ice thickness. Indeed, the monomers were able to stack to an extent as observed in negative stain micrographs (Figure 5-2c), and the number of stacked monomers can be controlled.

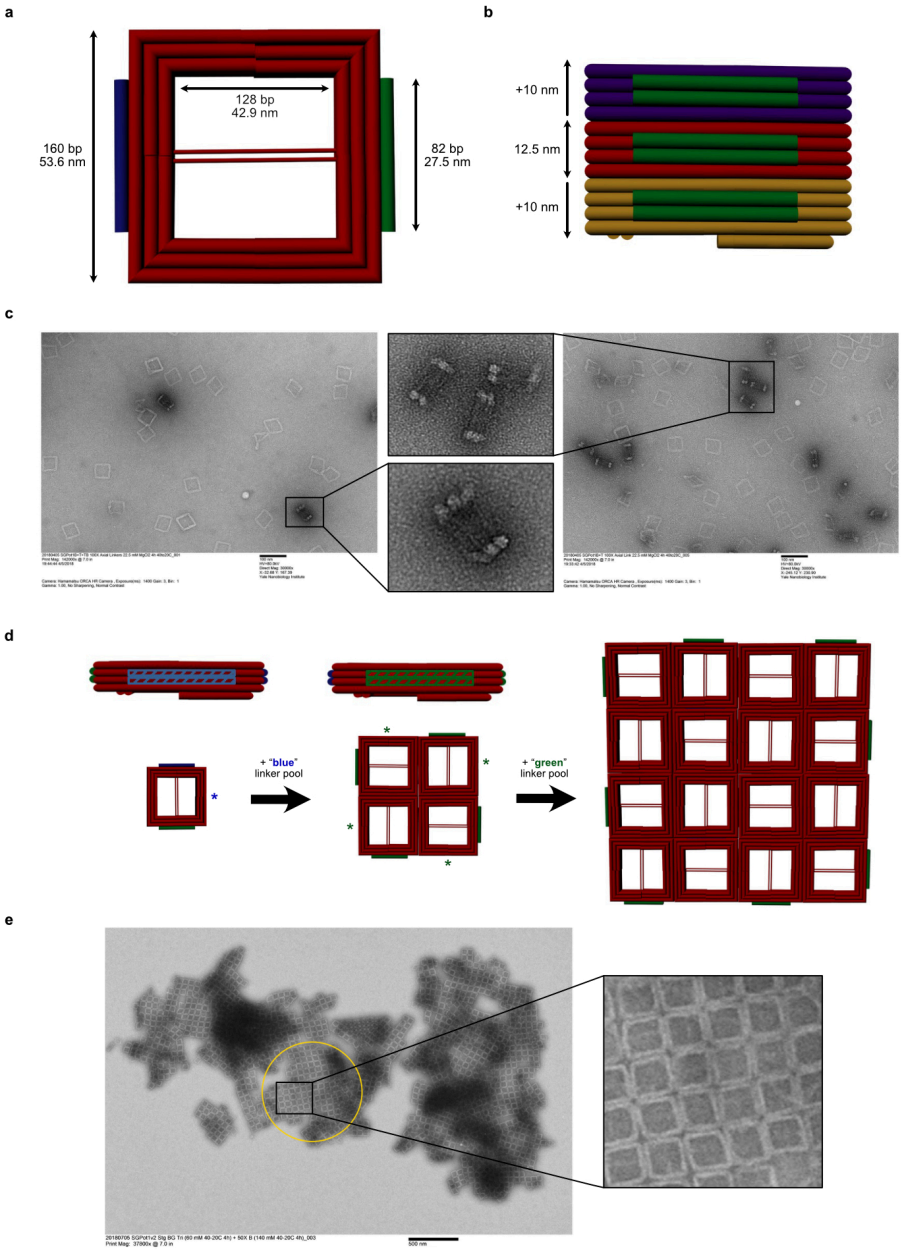

Figure 5-2. Pilot origami scaffold design. A) Top view of an origami monomer brick. Cylinders represent dsDNA helices, with two intervening ssDNA strands through the middle. Dimensions are shown. A 9-mer POT1 binding sequence was designed into one central strand for protein attachment. B) Monomers were designed to stack axially, in a modular manner, to precisely and systematically influence ice thickness. C) Negative stain micrographs demonstrating axial polymerization. D) Teeth (blue/green helices) are built into opposing sides and slot into corresponding grooves (blue/green dashed boxes) upon addition of specific oligonucleotide linkers, allowing us to laterally polymerize the monomer. Colored asterisks indicate the location of linker hybridization staple monomers together. E) Negative stain micrographs demonstrating lateral monolayer polymerization. The yellow circle indicates the size of a standard cryo-EM grid hole (1 μm). All 3D renderings were made by John Powell.

The sides of the scaffold were also sculpted such that addition of an oligonucleotide staple (Figure 5-2d, “blue” linkers) induces lateral tetramerization; the linker hybridizes to the origami, forming a tooth on one side and slots into a complementary groove on another monomer and joining them. The four-monomer unit is similarly incubated with another set of staples (“green” linkers) to stimulate total lateral polymerization. Our construct’s capacity to polymerize laterally was incredibly successful as judged by negative stain, showing continuous origami arrays with enough coverage to fill a standard cryo-EM grid (Figure 5-2e). With well-behaving origami constructs, we next focused on screening strategies for anchoring a POI to the origami scaffold.

5.2.2 Iterative protein-to-origami attachment strategies yield a robust anchor

The first anchoring strategy we tried involved a DNA-binding protein fusion with POT1 (protection of telomeres 1), a ssDNA-binding protein reportedly involved in telomeric maintenance and possessing a low nM ssDNA affinity[142]. In order to make the scaffold more widely adaptable, we also utilized SpyCatcher/SpyTag-based (SpyC/SpyT) covalent protein attachment[143]. SpyC/SpyT are two bioengineered components of a bacterial adhesin that spontaneously react in close proximity to form an isopeptide bond and covalently link both components. Instead of requiring a recombinantly-encoded POT1 fusion, separate purification of a POT1-SpyC chimera (19.6 + 15 kDa) then offers a smaller fusion tag to be encoded (SpyT, 13 residues without requirement for N-or C-terminal exposure) and has the advantage of a more universal tethering mechanism.

Accordingly, soluble hexamers incorporating CA-SpyT-MBP were assembled. Singly-tagged hexamers (Hex-SpyT-MBP) were isolated to ensure 1:1 binding with POT1-SpyC and therefore 1:1 stoichiometry with the scaffold tether site (Figure 5-2d). MBP was not cleaved so as to protect SpyT from degradation. In an attempt to purify a reacted complex of POT1-SpyC and Hex-SpyT-MBP however no interaction could be confirmed by gel filtration, even for hexamers with 6 incorporated SpyTs and MBP cleaved off (data not shown). Altogether the POT1 binding efficiency was extremely low (~10%; defined as the number of origami monomers with a single loaded complex) and caused us to pursue other anchoring strategies.

We also explored, and ultimately decided to use, oligonucleotide hybridization as an anchoring mechanism. We initially elected to take advantage of biotin-streptavidin (SA) interactions, one of the strongest affinities observed in biology (K_d ~10 fM for a SA tetramer[144]). We used a double mutant, S52G/R53D, with a 10-fold slower biotin off-rate[145]. Our designs incorporated a SA-SpyC fusion, intended to connect a biotinylated oligo with a POI containing SpyT (Figure 5-3). Ultimately however, ~30-50% of origami monomers showed successful loading of tetrameric SA alone, without POI tethering. Our impressions of the large SA tetramer size (~265 kDa particle) along with the number of intermediate connections – scaffold to oligo, oligo to SA, and SA to POI – did not justify further optimization.

We investigated a final tethering scheme involving POI fusion to a SNAP tag. A SNAP tag is a 20 kDa bioengineered alkyltransferase highly specific for

benzylguanine (BG)-modified oligonucleotides that forms a rapid and irreversible covalent attachment under physiological conditions[146, 147] (Figure 5-3c). We cloned and purified CA constructs containing a SNAP tag in place of SpyT, then cleaved off MBP and assembled single SNAP-tagged hexamers as before. The reaction with BG-oligonucleotides and CA-SNAP constructs was highly efficient, and incubation of the origami scaffold v3 with CA-SNAP-[BG]-oligo resulted in loading efficiencies much higher than observed for SpyC/SpyT-based protein engineering. In this background, the vast majority of origami monomers contained at least 1 and up to 4 hexamers loaded (Figure 5-3d).

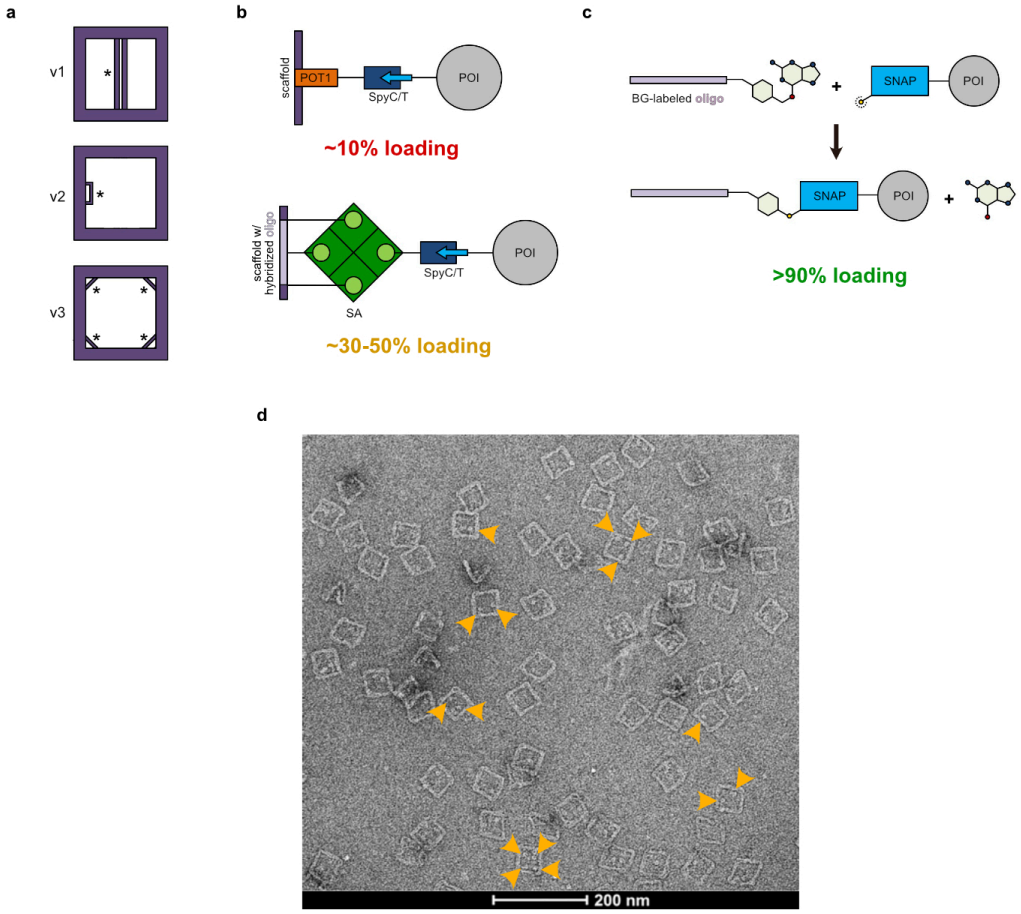

Figure 5-3. Summary of primary strategies used for origami loading. A) Iterative designs of the origami monomer that optimized for increased conformational sampling space to generate more particle orientations (v2) and the number of attachment sites (or "handles", shown by asterisks) to maximize particle capacity per origami monomer (v3). B) Screened methods of anchoring a POI to the origami scaffold including the DNA-binding protein POT1 (top) and a biotinylated oligonucleotide that binds streptavidin (SA) and is complimentary to programmed sites in the scaffold. Both anchors are tethered to the POI using SpyC/SpyT linkages in these constructs. C) Final method for anchoring a POI using a SNAP tag fusion, which covalently attaches to BG-labeled oligonucleotides and produces free guanine. D) Representative negative stain micrograph of a loading reaction using the anchor strategy in panel C. Singly SNAP-tagged CA hexamers conjugated to a BG oligo were loaded onto scaffold v3, which contains four handles. Orange arrows indicate select loaded origami monomers.

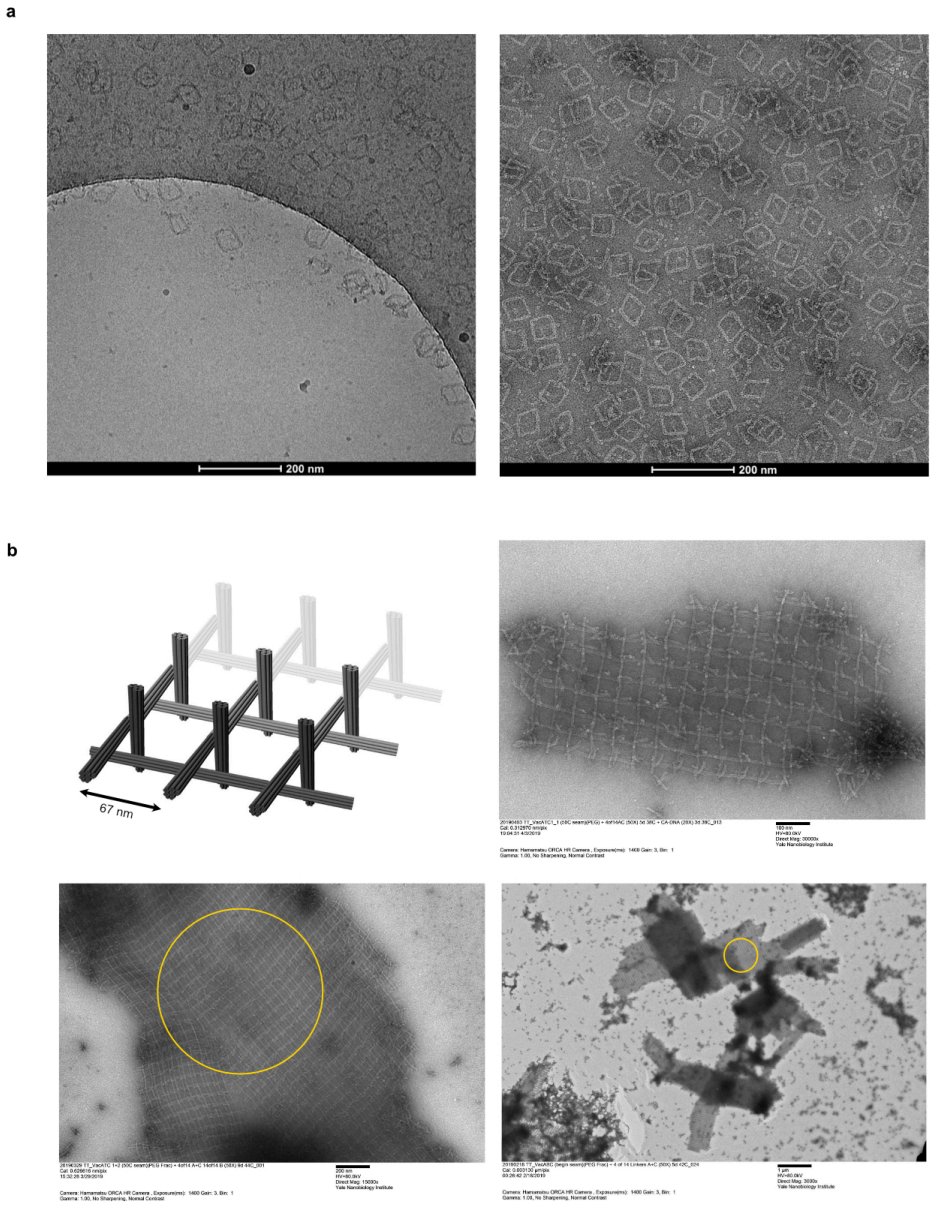

Figure 5-4. Cryo-EM grid behavior of our prototype origami scaffold and future designs. A) Examples of both “apo” (left) and loaded origami (right) exhibit a marked preference for the grid support structure instead of the grid hole. B) Rendering of tensegrity triangles which became the next fundamental scaffold design, with three representative images showing its capacity for multi-layered polymerization. The yellow circle indicates the size of a standard cryo-EM grid hole (1 μm). 3D rendering was made by John Powell.

5.3 Discussion

In our pursuit to mitigate the AWI problem, we aimed to systematically optimize our scaffold for loading efficiency. Our origami exhibited robust polymerization capability. Our screens showed a clearly superior attachment strategy using BG-oligonucleotides and a SNAP-tagged POI. Recent data suggest POT1 DNA affinity may be up to 1000-fold weaker than previously reported[148], and we didn't anticipate that the uranyl acetate salt used for staining (and $MgCl_2$ concentrations necessary for origami stability) might out-compete POT1 binding from the charged ssDNA in our screens. This likely explains our low binding POT1-mediated loading efficiency. We decided to explore oligonucleotide-based loading, which serves a few advantages: each oligonucleotide can be sequence-specifically targeted to handles within the scaffold without difficulty. Oligos also offer an array of chemistry for linkage to other macromolecules, and salt-sensitivity is a minor concern since DNA duplex stability is largely driven by the hydrophobic effect. The SA tag we screened, which binds biotinylated handles, was more efficient than POT1 but far from the efficiency required. Since the tag alone exhibited suboptimal loading efficiency, we reasoned that attachment to a POI could only be worse. Ultimately, SNAP-tagging methods were superior in loading our hexamers and was the chosen anchoring method.

With a robust tethering strategy and polymeric architecture, we attempted to analyze our scaffold's practicality for sequestering the CA hexamer from the AWI and allowing the hexamer to sample multiple orientations. However, we immediately struggled with origami behavior on cryo-EM grids. Our origami

constructs showed a marked preference for the grid support structure and avoided entering the grid hole (Figure 5-4a). In addition, the scaffold itself is ~5 nm thick on all sides and altogether renders ~30% of the origami's 2D surface area unusable for data collection. The DNA thickness causes a tradeoff in efficient use of grid space, lowering particle density and potentially contributing to higher background signal due to significant amounts of electron-rich DNA.

5.4 Future Directions

My role in the project was sunlit as I moved into work with SAMHD1 and CA, but further scaffolds were evaluated. These include support structures made with long, rigid viral filament proteins from members of the *Podoviridae* family, an amphipathic bacterial protein BslA that self-assembles into biofilms at the AWI[149], and other polymeric origami structures as shown in Figure 5-4b. The latter case is assembled from "tensegrity triangle" subunits. Further work must be done to optimize the scaffold array's stability while minimizing occupied grid space.

5.5 Experimental Procedures

5.5.1 Purification and assembly of singly-tagged Hex-SpyT-MBP

$CA_{A14C/E45C/W184A/M185A}$ was expressed and purified as described in Section 4.5.1. $CA_{A14C/E45C/W184A/M185A}$-SpyT-[mPro]-MBP-6xHis was expressed and purified identically to $CA_{A14C/E45C}$-[mPro]-MBP-6xHis as also described in Section 4.5.1, and hexamers were assembled using an identical dialysis protocol with a 4:1 ratio of CA:CA-SpyT. Mixed hexamers were separated by SpyT multiplicity using a MonoQ gradient fractionation, as previously described[150].

5.5.2 Purification of tether proteins

6xHis-POT1-SpyC was cloned into pMAT9 (Amp) lacking an MBP tag. Protein was expressed in *E. coli* BL21 (DE3) cells and induced at OD_{600} ~ 0.8 for 16 hrs at 18 ˚C. Cells were resuspended and lysed in Ni_A buffer supplemented with a protease inhibitor cocktail tablet (Roche) using a microfluidizer (15k psi). 6xHis-POT1-SpyC was purified through gravity Ni-NTA affinity and S200 PG 16/60 gel filtration chromatrography, flash frozen in 50 mM Tris pH 8.0, 100 mM NaCl, 0.1 mM TCEP, and stored at -80 ˚C until use.

The streptavidin variant (S52G, R53D) we purchased (Kerafast) is known as Traptavidin, developed by the Howarth lab[145].

5.5.3 Oligonucleotide preparation and BG-oligo reaction

For streptavidin experiments, biotin-labeled oligonucleotides were ordered from IDT. For SNAP-tag based oligonucleotide labeling, 4 µL of 5AmMC6-modified oligo (IDT; 2 mM in water) was mixed with 8 µL HEPES pH 8.5 and 12 µL BG-GLA-NHS (New England Biolabs; 20 mM in DMSO) and incubated for 1 hr at room temperature. The BG-oligo reaction product was confirmed by gel shift on a 20% TBE agarose gel and the labeled oligo was stored at -80 ˚C until further use.

5.5.4 Origami scaffold loading

Folded, purified origami monomers were obtained from John Powell. Generally, purified hexamers incorporating CA-SpyT(-MBP) or CA-SNAP were incubated in 2x molar excess of available oligonucleotide binding sites (“handles”). Origami monomers containing 4 oligonucleotide handles were therefore mixed to a final concentration of 25 nM origami (=100 nM handles), ~200 nM hexamer, and

30 mM $MgCl_2$ to maintain origami stability. Samples were incubated at 37 ˚C for 2-3 hrs and imaged via negative stain EM.

5.5.5 Negative stain EM

Negative staining was performed as described in Section 3.5.3 or by the Lin lab. At least 2-3 micrographs per screening condition were obtained, representing dozens of origami particles for loading efficiency estimation. Loading efficiency is defined as the number of origami monomers with at least one tethered POI, divided by total number of monomers observed.

5.6 Contributions

John Powell was the origami production side of the project and source of folded origami that we used for loading reactions, and also was principally involved in DNA construct design. He collected many of the negative stain images shown here alongside our efforts, with additional help and expertise from Qi Shen, Kaifeng Zhou, and Chenxiang Lin. I would also like to thank Yinkai Duan, an intern from Tsinghua University in our lab who helped work on the SNAP tag cloning and screening.

References

1. Amie, S. M.; Noble, E.; Kim, B., Intracellular nucleotide levels and the control of retroviral infections. *Virology* **2013,** *436* (2), 247-54.
2. Bjursell, G.; Skoog, L., Control of nucleotide pools in mammalian cells. *Antibiot Chemother (1971)* **1980,** *28*, 78-85.
3. Kumar, D.; Abdulovic, A. L.; Viberg, J.; Nilsson, A. K.; Kunkel, T. A.; Chabes, A., Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Res* **2011,** *39* (4), 1360-71.
4. Coggins, S. A.; Mahboubi, B.; Schinazi, R. F.; Kim, B., Mechanistic cross-talk between DNA/RNA polymerase enzyme kinetics and nucleotide substrate availability in cells: Implications for polymerase inhibitor discovery. *J Biol Chem* **2020,** *295* (39), 13432-13443.
5. Yao, N. Y.; Schroeder, J. W.; Yurieva, O.; Simmons, L. A.; O'Donnell, M. E., Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc Natl Acad Sci U S A* **2013,** *110* (32), 12942-7.
6. Hofer, A.; Crona, M.; Logan, D. T.; Sjoberg, B. M., DNA building blocks: keeping control of manufacture. *Crit Rev Biochem Mol Biol* **2012,** *47* (1), 50-63.
7. Franzolin, E.; Pontarin, G.; Rampazzo, C.; Miazzi, C.; Ferraro, P.; Palumbo, E.; Reichard, P.; Bianchi, V., The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci U S A* **2013,** *110* (35), 14272-7.
8. Seto, D.; Bhatnagar, S. K.; Bessman, M. J., The purification and properties of deoxyguanosine triphosphate triphosphohydrolase from Escherichia coli. *J Biol Chem* **1988,** *263* (3), 1494-9.
9. Zhu, C. F.; Wei, W.; Peng, X.; Dong, Y. H.; Gong, Y.; Yu, X. F., The mechanism of substrate-controlled allosteric regulation of SAMHD1 activated by GTP. *Acta Crystallogr D Biol Crystallogr* **2015,** *71* (Pt 3), 516-24.
10. Ji, X.; Wu, Y.; Yan, J.; Mehrens, J.; Yang, H.; DeLucia, M.; Hao, C.; Gronenborn, A. M.; Skowronski, J.; Ahn, J.; Xiong, Y., Mechanism of allosteric activation of SAMHD1 by dGTP. *Nat Struct Mol Biol* **2013,** *20* (11), 1304-9.
11. Goldstone, D. C.; Ennis-Adeniran, V.; Hedden, J. J.; Groom, H. C.; Rice, G. I.; Christodoulou, E.; Walker, P. A.; Kelly, G.; Haire, L. F.; Yap, M. W.; de Carvalho, L. P.; Stoye, J. P.; Crow, Y. J.; Taylor, I. A.; Webb, M., HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **2011,** *480* (7377), 379-82.
12. Yan, J.; Kaur, S.; DeLucia, M.; Hao, C.; Mehrens, J.; Wang, C.; Golczak, M.; Palczewski, K.; Gronenborn, A. M.; Ahn, J.; Skowronski, J., Tetramerization of SAMHD1 is required for biological activity and inhibition of HIV infection. *J Biol Chem* **2013,** *288* (15), 10406-17.
13. DeLucia, M.; Mehrens, J.; Wu, Y.; Ahn, J., HIV-2 and SIVmac accessory virulence factor Vpx down-regulates SAMHD1 enzyme catalysis prior to proteasome-dependent degradation. *J Biol Chem* **2013,** *288* (26), 19116-26.

14. Amie, S. M.; Bambara, R. A.; Kim, B., GTP is the primary activator of the anti-HIV restriction factor SAMHD1. *J Biol Chem* **2013,** *288* (35), 25001-25006.
15. Jang, S.; Zhou, X.; Ahn, J., Substrate Specificity of SAMHD1 Triphosphohydrolase Activity Is Controlled by Deoxyribonucleoside Triphosphates and Phosphorylation at Thr592. *Biochemistry* **2016,** *55* (39), 5635-5646.
16. Zhu, C.; Gao, W.; Zhao, K.; Qin, X.; Zhang, Y.; Peng, X.; Zhang, L.; Dong, Y.; Zhang, W.; Li, P.; Wei, W.; Gong, Y.; Yu, X. F., Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. *Nat Commun* **2013,** *4*, 2722.
17. Ji, X.; Tang, C.; Zhao, Q.; Wang, W.; Xiong, Y., Structural basis of cellular dNTP regulation by SAMHD1. *Proc Natl Acad Sci U S A* **2014,** *111* (41), E4305-14.
18. Koharudin, L. M.; Wu, Y.; DeLucia, M.; Mehrens, J.; Gronenborn, A. M.; Ahn, J., Structural basis of allosteric activation of sterile alpha motif and histidine-aspartate domain-containing protein 1 (SAMHD1) by nucleoside triphosphates. *J Biol Chem* **2014,** *289* (47), 32617-27.
19. Aravind, L.; Koonin, E. V., The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends in Biochemical Sciences* **1998,** *23* (12), 469-472.
20. Diamond, T. L.; Roshal, M.; Jamburuthugoda, V. K.; Reynolds, H. M.; Merriam, A. R.; Lee, K. Y.; Balakrishnan, M.; Bambara, R. A.; Planelles, V.; Dewhurst, S.; Kim, B., Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* **2004,** *279* (49), 51545-53.
21. Kennedy, E. M.; Gavegnano, C.; Nguyen, L.; Slater, R.; Lucas, A.; Fromentin, E.; Schinazi, R. F.; Kim, B., Ribonucleoside triphosphates as substrate of human immunodeficiency virus type 1 reverse transcriptase in human macrophages. *J Biol Chem* **2010,** *285* (50), 39380-91.
22. Hansen, E. C.; Seamon, K. J.; Cravens, S. L.; Stivers, J. T., GTP activator and dNTP substrates of HIV-1 restriction factor SAMHD1 generate a long-lived activated state. *Proc Natl Acad Sci U S A* **2014,** *111* (18), E1843-51.
23. Wang, Z.; Bhattacharya, A.; Villacorta, J.; Diaz-Griffero, F.; Ivanov, D. N., Allosteric Activation of SAMHD1 Protein by Deoxynucleotide Triphosphate (dNTP)-dependent Tetramerization Requires dNTP Concentrations That Are Similar to dNTP Concentrations Observed in Cycling T Cells. *J Biol Chem* **2016,** *291* (41), 21407-21413.
24. Arnold, L. H.; Groom, H. C.; Kunzelmann, S.; Schwefel, D.; Caswell, S. J.; Ordonez, P.; Mann, M. C.; Rueschenbaum, S.; Goldstone, D. C.; Pennell, S.; Howell, S. A.; Stoye, J. P.; Webb, M.; Taylor, I. A.; Bishop, K. N., Phospho-dependent Regulation of SAMHD1 Oligomerisation Couples Catalysis and Restriction. *PLoS Pathog* **2015,** *11* (10), e1005194.
25. Schmidt, S.; Schenkova, K.; Adam, T.; Erikson, E.; Lehmann-Koch, J.; Sertel, S.; Verhasselt, B.; Fackler, O. T.; Lasitschka, F.; Keppler, O. T.,

SAMHD1's protein expression profile in humans. *J Leukoc Biol* **2015,** *98* (1), 5-14.
26. Coggins, S. A.; Mahboubi, B.; Schinazi, R. F.; Kim, B., SAMHD1 Functions and Human Diseases. *Viruses* **2020,** *12* (4).
27. Cribier, A.; Descours, B.; Valadao, A. L.; Laguette, N.; Benkirane, M., Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. *Cell Rep* **2013,** *3* (4), 1036-43.
28. Baldauf, H. M.; Pan, X.; Erikson, E.; Schmidt, S.; Daddacha, W.; Burggraf, M.; Schenkova, K.; Ambiel, I.; Wabnitz, G.; Gramberg, T.; Panitz, S.; Flory, E.; Landau, N. R.; Sertel, S.; Rutsch, F.; Lasitschka, F.; Kim, B.; Konig, R.; Fackler, O. T.; Keppler, O. T., SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nat Med* **2012,** *18* (11), 1682-7.
29. St Gelais, C.; de Silva, S.; Amie, S. M.; Coleman, C. M.; Hoy, H.; Hollenbaugh, J. A.; Kim, B.; Wu, L., SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. *Retrovirology* **2012,** *9*, 105.
30. Goujon, C.; Schaller, T.; Galao, R. P.; Amie, S. M.; Kim, B.; Olivieri, K.; Neil, S. J.; Malim, M. H., Evidence for IFNalpha-induced, SAMHD1-independent inhibitors of early HIV-1 infection. *Retrovirology* **2013,** *10*, 23.
31. Brandariz-Nunez, A.; Valle-Casuso, J. C.; White, T. E.; Nguyen, L.; Bhattacharya, A.; Wang, Z.; Demeler, B.; Amie, S.; Knowlton, C.; Kim, B.; Ivanov, D. N.; Diaz-Griffero, F., Contribution of oligomerization to the anti-HIV-1 properties of SAMHD1. *Retrovirology* **2013,** *10*, 131.
32. White, T. E.; Brandariz-Nunez, A.; Valle-Casuso, J. C.; Amie, S.; Nguyen, L.; Kim, B.; Brojatsch, J.; Diaz-Griffero, F., Contribution of SAM and HD domains to retroviral restriction mediated by human SAMHD1. *Virology* **2013,** *436* (1), 81-90.
33. Chen, Z.; Zhu, M.; Pan, X.; Zhu, Y.; Yan, H.; Jiang, T.; Shen, Y.; Dong, X.; Zheng, N.; Lu, J.; Ying, S.; Shen, Y., Inhibition of Hepatitis B virus replication by SAMHD1. *Biochem Biophys Res Commun* **2014,** *450* (4), 1462-8.
34. Qin, Z.; Bonifati, S.; St Gelais, C.; Li, T. W.; Kim, S. H.; Antonucci, J. M.; Mahboubi, B.; Yount, J. S.; Xiong, Y.; Kim, B.; Wu, L., The dNTPase activity of SAMHD1 is important for its suppression of innate immune responses in differentiated monocytic cells. *J Biol Chem* **2020,** *295* (6), 1575-1586.
35. Guo, H.; Wei, W.; Wei, Z.; Liu, X.; Evans, S. L.; Yang, W.; Wang, H.; Guo, Y.; Zhao, K.; Zhou, J. Y.; Yu, X. F., Identification of critical regions in human SAMHD1 required for nuclear localization and Vpx-mediated degradation. *PLoS One* **2013,** *8* (7), e66201.
36. Schott, K.; Fuchs, N. V.; Derua, R.; Mahboubi, B.; Schnellbacher, E.; Seifried, J.; Tondera, C.; Schmitz, H.; Shepard, C.; Brandariz-Nunez, A.; Diaz-Griffero, F.; Reuter, A.; Kim, B.; Janssens, V.; Konig, R., Dephosphorylation of the HIV-1 restriction factor SAMHD1 is mediated by PP2A-B55alpha holoenzymes during mitotic exit. *Nat Commun* **2018,** *9* (1), 2227.

37. Hu, J.; Gao, Q.; Yang, Y.; Xia, J.; Zhang, W.; Chen, Y.; Zhou, Z.; Chang, L.; Hu, Y.; Zhou, H.; Liang, L.; Li, X.; Long, Q.; Wang, K.; Huang, A.; Tang, N., Hexosamine biosynthetic pathway promotes the antiviral activity of SAMHD1 by enhancing O-GlcNAc transferase-mediated protein O-GlcNAcylation. *Theranostics* **2021,** *11* (2), 805-823.
38. Mauney, C. H.; Rogers, L. C.; Harris, R. S.; Daniel, L. W.; Devarie-Baez, N. O.; Wu, H.; Furdui, C. M.; Poole, L. B.; Perrino, F. W.; Hollis, T., The SAMHD1 dNTP Triphosphohydrolase Is Controlled by a Redox Switch. *Antioxid Redox Signal* **2017,** *27* (16), 1317-1331.
39. Wang, Z.; Bhattacharya, A.; White, T.; Buffone, C.; McCabe, A.; Nguyen, L. A.; Shepard, C. N.; Pardo, S.; Kim, B.; Weintraub, S. T.; Demeler, B.; Diaz-Griffero, F.; Ivanov, D. N., Functionality of Redox-Active Cysteines Is Required for Restriction of Retroviral Replication by SAMHD1. *Cell Rep* **2018,** *24* (4), 815-823.
40. Martinat, C.; Cormier, A.; Tobaly-Tapiero, J.; Palmic, N.; Casartelli, N.; Mahboubi, B.; Coggins, S. A.; Buchrieser, J.; Persaud, M.; Diaz-Griffero, F.; Espert, L.; Bossis, G.; Lesage, P.; Schwartz, O.; Kim, B.; Margottin-Goguet, F.; Saib, A.; Zamborlini, A., SUMOylation of SAMHD1 at Lysine 595 is required for HIV-1 restriction in non-cycling cells. *Nature Communications* **2021,** *12* (1).
41. Li, Z.; Huan, C.; Wang, H.; Liu, Y.; Liu, X.; Su, X.; Yu, J.; Zhao, Z.; Yu, X. F.; Zheng, B.; Zhang, W., TRIM21-mediated proteasomal degradation of SAMHD1 regulates its antiviral activity. *EMBO Rep* **2020,** *21* (1), e47528.
42. Yan, J.; Hao, C.; DeLucia, M.; Swanson, S.; Florens, L.; Washburn, M. P.; Ahn, J.; Skowronski, J., CyclinA2-Cyclin-dependent Kinase Regulates SAMHD1 Protein Phosphohydrolase Domain. *J Biol Chem* **2015,** *290* (21), 13279-92.
43. St Gelais, C.; Kim, S. H.; Maksimova, V. V.; Buzovetsky, O.; Knecht, K. M.; Shepard, C.; Kim, B.; Xiong, Y.; Wu, L., A Cyclin-Binding Motif in Human SAMHD1 Is Required for Its HIV-1 Restriction, dNTPase Activity, Tetramer Formation, and Efficient Phosphorylation. *J Virol* **2018,** *92* (6).
44. Tang, C.; Ji, X.; Wu, L.; Xiong, Y., Impaired dNTPase activity of SAMHD1 by phosphomimetic mutation of Thr-592. *J Biol Chem* **2015,** *290* (44), 26352-9.
45. Bhattacharya, A.; Wang, Z.; White, T.; Buffone, C.; Nguyen, L. A.; Shepard, C. N.; Kim, B.; Demeler, B.; Diaz-Griffero, F.; Ivanov, D. N., Effects of T592 phosphomimetic mutations on tetramer stability and dNTPase activity of SAMHD1 can not explain the retroviral restriction defect. *Sci Rep* **2016,** *6*, 31353.
46. Batalis, S.; Rogers, L. C.; Hemphill, W. O.; Mauney, C. H.; Ornelles, D. A.; Hollis, T., SAMHD1 Phosphorylation at T592 Regulates Cellular Localization and S-phase Progression. *Front Mol Biosci* **2021,** *8*, 724870.
47. Crow, Y. J.; Chase, D. S.; Lowenstein Schmidt, J.; Szynkiewicz, M.; Forte, G. M.; Gornall, H. L.; Oojageer, A.; Anderson, B.; Pizzino, A.; Helman, G.; Abdel-Hamid, M. S.; Abdel-Salam, G. M.; Ackroyd, S.; Aeby, A.; Agosta, G.; Albin, C.; Allon-Shalev, S.; Arellano, M.; Ariaudo, G.; Aswani, V.;

Babul-Hirji, R.; Baildam, E. M.; Bahi-Buisson, N.; Bailey, K. M.; Barnerias, C.; Barth, M.; Battini, R.; Beresford, M. W.; Bernard, G.; Bianchi, M.; Billette de Villemeur, T.; Blair, E. M.; Bloom, M.; Burlina, A. B.; Carpanelli, M. L.; Carvalho, D. R.; Castro-Gago, M.; Cavallini, A.; Cereda, C.; Chandler, K. E.; Chitayat, D. A.; Collins, A. E.; Sierra Corcoles, C.; Cordeiro, N. J.; Crichiutti, G.; Dabydeen, L.; Dale, R. C.; D'Arrigo, S.; De Goede, C. G.; De Laet, C.; De Waele, L. M.; Denzler, I.; Desguerre, I.; Devriendt, K.; Di Rocco, M.; Fahey, M. C.; Fazzi, E.; Ferrie, C. D.; Figueiredo, A.; Gener, B.; Goizet, C.; Gowrinathan, N. R.; Gowrishankar, K.; Hanrahan, D.; Isidor, B.; Kara, B.; Khan, N.; King, M. D.; Kirk, E. P.; Kumar, R.; Lagae, L.; Landrieu, P.; Lauffer, H.; Laugel, V.; La Piana, R.; Lim, M. J.; Lin, J. P.; Linnankivi, T.; Mackay, M. T.; Marom, D. R.; Marques Lourenco, C.; McKee, S. A.; Moroni, I.; Morton, J. E.; Moutard, M. L.; Murray, K.; Nabbout, R.; Nampoothiri, S.; Nunez-Enamorado, N.; Oades, P. J.; Olivieri, I.; Ostergaard, J. R.; Perez-Duenas, B.; Prendiville, J. S.; Ramesh, V.; Rasmussen, M.; Regal, L.; Ricci, F.; Rio, M.; Rodriguez, D.; Roubertie, A.; Salvatici, E.; Segers, K. A.; Sinha, G. P.; Soler, D.; Spiegel, R.; Stodberg, T. I.; Straussberg, R.; Swoboda, K. J.; Suri, M.; Tacke, U.; Tan, T. Y.; te Water Naude, J.; Wee Teik, K.; Thomas, M. M.; Till, M.; Tonduti, D.; Valente, E. M.; Van Coster, R. N.; van der Knaap, M. S.; Vassallo, G.; Vijzelaar, R.; Vogt, J.; Wallace, G. B.; Wassmer, E.; Webb, H. J.; Whitehouse, W. P.; Whitney, R. N.; Zaki, M. S.; Zuberi, S. M.; Livingston, J. H.; Rozenberg, F.; Lebon, P.; Vanderver, A.; Orcesi, S.; Rice, G. I., Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. *Am J Med Genet A* **2015,** *167A* (2), 296-312.

48. Rice, G. I.; Bond, J.; Asipu, A.; Brunette, R. L.; Manfield, I. W.; Carr, I. M.; Fuller, J. C.; Jackson, R. M.; Lamb, T.; Briggs, T. A.; Ali, M.; Gornall, H.; Couthard, L. R.; Aeby, A.; Attard-Montalto, S. P.; Bertini, E.; Bodemer, C.; Brockmann, K.; Brueton, L. A.; Corry, P. C.; Desguerre, I.; Fazzi, E.; Cazorla, A. G.; Gener, B.; Hamel, B. C.; Heiberg, A.; Hunter, M.; van der Knaap, M. S.; Kumar, R.; Lagae, L.; Landrieu, P. G.; Lourenco, C. M.; Marom, D.; McDermott, M. F.; van der Merwe, W.; Orcesi, S.; Prendiville, J. S.; Rasmussen, M.; Shalev, S. A.; Soler, D. M.; Shinawi, M.; Spiegel, R.; Tan, T. Y.; Vanderver, A.; Wakeling, E. L.; Wassmer, E.; Whittaker, E.; Lebon, P.; Stetson, D. B.; Bonthron, D. T.; Crow, Y. J., Mutations involved in Aicardi-Goutieres syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat Genet* **2009,** *41* (7), 829-32.
49. Dharan, A.; Bachmann, N.; Talley, S.; Zwikelmaier, V.; Campbell, E. M., Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nat Microbiol* **2020,** *5* (9), 1088-1095.
50. Seamon, K. J.; Sun, Z.; Shlyakhtenko, L. S.; Lyubchenko, Y. L.; Stivers, J. T., SAMHD1 is a single-stranded nucleic acid binding protein with no active site-associated nuclease activity. *Nucleic Acids Res* **2015,** *43* (13), 6486-99.
51. Yu, C. H.; Bhattacharya, A.; Persaud, M.; Taylor, A. B.; Wang, Z.; Bulnes-Ramos, A.; Xu, J.; Selyutina, A.; Martinez-Lopez, A.; Cano, K.; Demeler,

B.; Kim, B.; Hardies, S. C.; Diaz-Griffero, F.; Ivanov, D. N., Nucleic acid binding by SAMHD1 contributes to the antiretroviral activity and is enhanced by the GpsN modification. *Nat Commun* **2021,** *12* (1), 731.
52. Antonucci, J. M.; Kim, S. H.; St Gelais, C.; Bonifati, S.; Li, T. W.; Buzovetsky, O.; Knecht, K. M.; Duchon, A. A.; Xiong, Y.; Musier-Forsyth, K.; Wu, L., SAMHD1 Impairs HIV-1 Gene Expression and Negatively Modulates Reactivation of Viral Latency in CD4(+) T Cells. *J Virol* **2018,** *92* (15).
53. Zhao, K.; Du, J.; Han, X.; Goodier, J. L.; Li, P.; Zhou, X.; Wei, W.; Evans, S. L.; Li, L.; Zhang, W.; Cheung, L. E.; Wang, G.; Kazazian, H. H., Jr.; Yu, X. F., Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep* **2013,** *4* (6), 1108-15.
54. White, T. E.; Brandariz-Nunez, A.; Martinez-Lopez, A.; Knowlton, C.; Lenzi, G.; Kim, B.; Ivanov, D.; Diaz-Griffero, F., A SAMHD1 mutation associated with Aicardi-Goutieres syndrome uncouples the ability of SAMHD1 to restrict HIV-1 from its ability to downmodulate type I interferon in humans. *Hum Mutat* **2017,** *38* (6), 658-668.
55. Hofmann, H.; Logue, E. C.; Bloch, N.; Daddacha, W.; Polsky, S. B.; Schultz, M. L.; Kim, B.; Landau, N. R., The Vpx lentiviral accessory protein targets SAMHD1 for degradation in the nucleus. *J Virol* **2012,** *86* (23), 12552-60.
56. Spragg, C. J.; Emerman, M., Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proc Natl Acad Sci U S A* **2013,** *110* (52), 21136-41.
57. Skasko, M.; Weiss, K. K.; Reynolds, H. M.; Jamburuthugoda, V.; Lee, K.; Kim, B., Mechanistic differences in RNA-dependent DNA polymerization and fidelity between murine leukemia virus and HIV-1 reverse transcriptases. *J Biol Chem* **2005,** *280* (13), 12190-200.
58. Sunseri, N.; O'Brien, M.; Bhardwaj, N.; Landau, N. R., Human immunodeficiency virus type 1 modified to package Simian immunodeficiency virus Vpx efficiently infects macrophages and dendritic cells. *J Virol* **2011,** *85* (13), 6263-74.
59. Gao, D.; Wu, J.; Wu, Y. T.; Du, F.; Aroh, C.; Yan, N.; Sun, L.; Chen, Z. J., Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **2013,** *341* (6148), 903-6.
60. Zhang, K.; Lv, D. W.; Li, R., Conserved Herpesvirus Protein Kinases Target SAMHD1 to Facilitate Virus Replication. *Cell Rep* **2019,** *28* (2), 449-459 e5.
61. Monit, C.; Morris, E. R.; Ruis, C.; Szafran, B.; Thiltgen, G.; Tsai, M. C.; Mitchison, N. A.; Bishop, K. N.; Stoye, J. P.; Taylor, I. A.; Fassati, A.; Goldstein, R. A., Positive selection in dNTPase SAMHD1 throughout mammalian evolution. *Proc Natl Acad Sci U S A* **2019,** *116* (37), 18647-18654.
62. Nakai, H.; Richardson, C. C., The gene 1.2 protein of bacteriophage T7 interacts with the Escherichia coli dGTP triphosphohydrolase to form a GTP-binding protein. *J Biol Chem* **1990,** *265* (8), 4411-9.

63. White, T. E.; Brandariz-Nunez, A.; Valle-Casuso, J. C.; Amie, S.; Nguyen, L. A.; Kim, B.; Tuzova, M.; Diaz-Griffero, F., The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation. *Cell Host Microbe* **2013,** *13* (4), 441-51.
64. Lahouassa, H.; Daddacha, W.; Hofmann, H.; Ayinde, D.; Logue, E. C.; Dragin, L.; Bloch, N.; Maudet, C.; Bertrand, M.; Gramberg, T.; Pancino, G.; Priet, S.; Canard, B.; Laguette, N.; Benkirane, M.; Transy, C.; Landau, N. R.; Kim, B.; Margottin-Goguet, F., SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* **2012,** *13* (3), 223-228.
65. Kim, B.; Nguyen, L. A.; Daddacha, W.; Hollenbaugh, J. A., Tight interplay among SAMHD1 protein level, cellular dNTP levels, and HIV-1 proviral DNA synthesis kinetics in human primary monocyte-derived macrophages. *J Biol Chem* **2012,** *287* (26), 21570-4.
66. Hollenbaugh, J. A.; Tao, S.; Lenzi, G. M.; Ryu, S.; Kim, D. H.; Diaz-Griffero, F.; Schinazi, R. F.; Kim, B., dNTP pool modulation dynamics by SAMHD1 protein in monocyte-derived macrophages. *Retrovirology* **2014,** *11*, 63.
67. Rodrigues, V.; Ruffin, N.; San-Roman, M.; Benaroch, P., Myeloid Cell Interaction with HIV: A Complex Relationship. *Front Immunol* **2017,** *8*, 1698.
68. Tan, J.; Sattentau, Q. J., The HIV-1-containing macrophage compartment: a perfect cellular niche? *Trends Microbiol* **2013,** *21* (8), 405-12.
69. Szaniawski, M. A.; Spivak, A. M.; Bosque, A.; Planelles, V., Sex Influences SAMHD1 Activity and Susceptibility to Human Immunodeficiency Virus-1 in Primary Human Macrophages. *J Infect Dis* **2019,** *219* (5), 777-785.
70. Traut, T. W., Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem* **1994,** *140* (1), 1-22.
71. Herold, N.; Rudd, S. G.; Sanjiv, K.; Kutzner, J.; Myrberg, I. H.; Paulin, C. B. J.; Olsen, T. K.; Helleday, T.; Henter, J. I.; Schaller, T., With me or against me: Tumor suppressor and drug resistance activities of SAMHD1. *Exp Hematol* **2017,** *52*, 32-39.
72. Schaller, T.; Herold, N., Evidence for SAMHD1 Tumor Suppressor Functions in Acute Myeloid Leukemia. *Acta Haematol* **2020,** *143* (1), 7-8.
73. Daddacha, W.; Koyen, A. E.; Bastien, A. J.; Head, P. E.; Dhere, V. R.; Nabeta, G. N.; Connolly, E. C.; Werner, E.; Madden, M. Z.; Daly, M. B.; Minten, E. V.; Whelan, D. R.; Schlafstein, A. J.; Zhang, H.; Anand, R.; Doronio, C.; Withers, A. E.; Shepard, C.; Sundaram, R. K.; Deng, X.; Dynan, W. S.; Wang, Y.; Bindra, R. S.; Cejka, P.; Rothenberg, E.; Doetsch, P. W.; Kim, B.; Yu, D. S., SAMHD1 Promotes DNA End Resection to Facilitate DNA Repair by Homologous Recombination. *Cell Rep* **2017,** *20* (8), 1921-1935.
74. Coquel, F.; Silva, M. J.; Techer, H.; Zadorozhny, K.; Sharma, S.; Nieminuszczy, J.; Mettling, C.; Dardillac, E.; Barthe, A.; Schmitz, A. L.; Promonet, A.; Cribier, A.; Sarrazin, A.; Niedzwiedz, W.; Lopez, B.; Costanzo, V.; Krejci, L.; Chabes, A.; Benkirane, M.; Lin, Y. L.; Pasero, P., SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature* **2018,** *557* (7703), 57-61.

75. Majerska, J.; Feretzaki, M.; Glousker, G.; Lingner, J., Transformation-induced stress at telomeres is counteracted through changes in the telomeric proteome including SAMHD1. *Life Sci Alliance* **2018,** *1* (4), e201800121.
76. Clifford, R.; Louis, T.; Robbe, P.; Ackroyd, S.; Burns, A.; Timbs, A. T.; Wright Colopy, G.; Dreau, H.; Sigaux, F.; Judde, J. G.; Rotger, M.; Telenti, A.; Lin, Y. L.; Pasero, P.; Maelfait, J.; Titsias, M.; Cohen, D. R.; Henderson, S. J.; Ross, M. T.; Bentley, D.; Hillmen, P.; Pettitt, A.; Rehwinkel, J.; Knight, S. J.; Taylor, J. C.; Crow, Y. J.; Benkirane, M.; Schuh, A., SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood* **2014,** *123* (7), 1021-31.
77. Amin, N. A.; Seymour, E.; Saiya-Cork, K.; Parkin, B.; Shedden, K.; Malek, S. N., A Quantitative Analysis of Subclonal and Clonal Gene Mutations before and after Therapy in Chronic Lymphocytic Leukemia. *Clin Cancer Res* **2016,** *22* (17), 4525-35.
78. Wang, J. L.; Lu, F. Z.; Shen, X. Y.; Wu, Y.; Zhao, L. T., SAMHD1 is down regulated in lung cancer by methylation and inhibits tumor cell proliferation. *Biochem Biophys Res Commun* **2014,** *455* (3-4), 229-33.
79. de Silva, S.; Wang, F.; Hake, T. S.; Porcu, P.; Wong, H. K.; Wu, L., Downregulation of SAMHD1 expression correlates with promoter DNA methylation in Sezary syndrome patients. *J Invest Dermatol* **2014,** *134* (2), 562-565.
80. Merati, M.; Buethe, D. J.; Cooper, K. D.; Honda, K. S.; Wang, H.; Gerstenblith, M. R., Aggressive CD8(+) epidermotropic cutaneous T-cell lymphoma associated with homozygous mutation in SAMHD1. *JAAD Case Rep* **2015,** *1* (4), 227-9.
81. Jiang, H.; Li, C.; Liu, Z.; Shengjing, H.; Hu, Expression and Relationship of SAMHD1 with Other Apoptotic and Autophagic Genes in Acute Myeloid Leukemia Patients. *Acta Haematol* **2020,** *143* (1), 51-59.
82. Rentoft, M.; Lindell, K.; Tran, P.; Chabes, A. L.; Buckland, R. J.; Watt, D. L.; Marjavaara, L.; Nilsson, A. K.; Melin, B.; Trygg, J.; Johansson, E.; Chabes, A., Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proc Natl Acad Sci U S A* **2016,** *113* (17), 4723-8.
83. Johansson, P.; Klein-Hitpass, L.; Choidas, A.; Habenberger, P.; Mahboubi, B.; Kim, B.; Bergmann, A.; Scholtysik, R.; Brauser, M.; Lollies, A.; Siebert, R.; Zenz, T.; Duhrsen, U.; Kuppers, R.; Durig, J., SAMHD1 is recurrently mutated in T-cell prolymphocytic leukemia. *Blood Cancer J* **2018,** *8* (1), 11.
84. Chen, S.; Bonifati, S.; Qin, Z.; St Gelais, C.; Kodigepalli, K. M.; Barrett, B. S.; Kim, S. H.; Antonucci, J. M.; Ladner, K. J.; Buzovetsky, O.; Knecht, K. M.; Xiong, Y.; Yount, J. S.; Guttridge, D. C.; Santiago, M. L.; Wu, L., SAMHD1 suppresses innate immune responses to viral infections and inflammatory stimuli by inhibiting the NF-kappaB and interferon pathways. *Proc Natl Acad Sci U S A* **2018,** *115* (16), E3798-E3807.
85. Kim, E. T.; Roche, K. L.; Kulej, K.; Spruce, L. A.; Seeholzer, S. H.; Coen, D. M.; Diaz-Griffero, F.; Murphy, E. A.; Weitzman, M. D., SAMHD1

Modulates Early Steps during Human Cytomegalovirus Infection by Limiting NF-kappaB Activation. *Cell Rep* **2019,** *28* (2), 434-448 e6.
86. Oo, A.; Zandi, K.; Shepard, C.; Bassit, L. C.; Musall, K. L.; Kim, D.-H.; Schinazi, R. F.; Kim, B., Cellular Innate Immunity Activation Induced by Genetic Loss of Aicardi Goutieres Syndrome Protein Suppresses SARS-CoV-2 Replication. *SSRN Electronic Journal* **2020**.
87. Wichit, S.; Hamel, R.; Zanzoni, A.; Diop, F.; Cribier, A.; Talignani, L.; Diack, A.; Ferraris, P.; Liegeois, F.; Urbach, S.; Ekchariyawat, P.; Merits, A.; Yssel, H.; Benkirane, M.; Misse, D., SAMHD1 Enhances Chikungunya and Zika Virus Replication in Human Skin Fibroblasts. *Int J Mol Sci* **2019,** *20* (7).
88. Buzovetsky, O.; Tang, C.; Knecht, K. M.; Antonucci, J. M.; Wu, L.; Ji, X.; Xiong, Y., The SAM domain of mouse SAMHD1 is critical for its activation and regulation. *Nat Commun* **2018,** *9* (1), 411.
89. Guieze, R.; Robbe, P.; Clifford, R.; de Guibert, S.; Pereira, B.; Timbs, A.; Dilhuydy, M. S.; Cabes, M.; Ysebaert, L.; Burns, A.; Nguyen-Khac, F.; Davi, F.; Veronese, L.; Combes, P.; Le Garff-Tavernier, M.; Leblond, V.; Merle-Beral, H.; Alsolami, R.; Hamblin, A.; Mason, J.; Pettitt, A.; Hillmen, P.; Taylor, J.; Knight, S. J.; Tournilhac, O.; Schuh, A., Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood* **2015,** *126* (18), 2110-7.
90. Giannakis, M.; Mu, X. J.; Shukla, S. A.; Qian, Z. R.; Cohen, O.; Nishihara, R.; Bahl, S.; Cao, Y.; Amin-Mansour, A.; Yamauchi, M.; Sukawa, Y.; Stewart, C.; Rosenberg, M.; Mima, K.; Inamura, K.; Nosho, K.; Nowak, J. A.; Lawrence, M. S.; Giovannucci, E. L.; Chan, A. T.; Ng, K.; Meyerhardt, J. A.; Van Allen, E. M.; Getz, G.; Gabriel, S. B.; Lander, E. S.; Wu, C. J.; Fuchs, C. S.; Ogino, S.; Garraway, L. A., Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* **2016,** *15* (4), 857-865.
91. Bowen, N. E.; Temple, J.; Shepard, C.; Oo, A.; Arizaga, F.; Kapoor-Vazirani, P.; Persaud, M.; Yu, C. H.; Kim, D. H.; Schinazi, R. F.; Ivanov, D. N.; Diaz-Griffero, F.; Yu, D. S.; Xiong, Y.; Kim, B., Structural and functional characterization explains loss of dNTPase activity of the cancer-specific R366C/H mutant SAMHD1 proteins. *J Biol Chem* **2021**, 101170.
92. Hawe, A.; Sutter, M.; Jiskoot, W., Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **2008,** *25* (7), 1487-99.
93. Schwefel, D.; Groom, H. C.; Boucherit, V. C.; Christodoulou, E.; Walker, P. A.; Stoye, J. P.; Bishop, K. N.; Taylor, I. A., Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **2014,** *505* (7482), 234-8.
94. Ryoo, J.; Choi, J.; Oh, C.; Kim, S.; Seo, M.; Kim, S. Y.; Seo, D.; Kim, J.; White, T. E.; Brandariz-Nunez, A.; Diaz-Griffero, F.; Yun, C. H.; Hollenbaugh, J. A.; Kim, B.; Baek, D.; Ahn, K., The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat Med* **2014,** *20* (8), 936-41.
95. Seamon, K. J.; Bumpus, N. N.; Stivers, J. T., Single-Stranded Nucleic Acids Bind to the Tetramer Interface of SAMHD1 and Prevent Formation of the Catalytic Homotetramer. *Biochemistry* **2016,** *55* (44), 6087-6099.

96. Otwinowski, Z.; Minor, W., Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **1997,** *276*, 307-326.
97. McCoy, A.; Grosse-Kunstleve, R.; Adams, P.; Winn, M.; Storoni, L.; Read, R., Phaser crystallographic software. *J. Appl. Cryst.* **2007,** *40*, 658-674.
98. Murshudov, G.; Skubák, P.; Lebedev, A.; Pannu, N.; Steiner, R.; Nichols, R.; Winn, M.; Long, F.; Vagin, A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011,** *67*, 355-367.
99. Emsley, P.; Lohkamp, B.; Scott, W.; Cowtain, K., Features and development of Coot. . *Acta. Crystallogr. D Biol. Crystallogr.* **2010,** *66*, 486-501.
100. Wang, F.; St Gelais, C.; de Silva, S.; Zhang, H.; Geng, Y.; Shepard, C.; Kim, B.; Yount, J. S.; Wu, L., Phosphorylation of mouse SAMHD1 regulates its restriction of human immunodeficiency virus type 1 infection, but not murine leukemia virus infection. *Virology* **2016,** *487*, 273-84.
101. Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R., The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **2019,** *47* (W1), W636-W641.
102. Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K., The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **2004,** *32* (3), 1037-49.
103. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **2014,** *42* (Web Server issue), W320-4.
104. Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A., cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **2017,** *14* (3), 290-296.
105. Xu, C.; Fischer, D. K.; Rankovic, S.; Li, W.; Dick, R.; Runge, B.; Zadorozhnyi, R.; Ahn, J.; Aiken, C.; Polenova, T.; Engelman, A. N.; Ambrose, Z.; Rousso, I.; Perilla, J. R., Permeability of the HIV-1 capsid to metabolites modulates viral DNA synthesis. *bioRxiv* **2020**, 2020.04.30.071217.
106. Temple, J.; Tripler, T. N.; Shen, Q.; Xiong, Y., A snapshot of HIV-1 capsid-host interactions. *Curr Res Struct Biol* **2020,** *2*, 222-228.
107. Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M., X-ray structures of the hexameric building block of the HIV capsid. *Cell* **2009,** *137* (7), 1282-92.
108. Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **2013,** *497* (7451), 643-6.
109. Jacques, D. A.; McEwan, W. A.; Hilditch, L.; Price, A. J.; Towers, G. J.; James, L. C., HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature* **2016,** *536* (7616), 349-53.
110. Mallery, D. L.; Marquez, C. L.; McEwan, W. A.; Dickson, C. F.; Jacques, D. A.; Anandapadamanaban, M.; Bichel, K.; Towers, G. J.; Saiardi, A.;

Bocking, T.; James, L. C., IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife* **2018,** *7*.
111. Malikov, V.; da Silva, E. S.; Jovasevic, V.; Bennett, G.; de Souza Aranha Vieira, D. A.; Schulte, B.; Diaz-Griffero, F.; Walsh, D.; Naghavi, M. H., HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus. *Nat Commun* **2015,** *6*, 6660.
112. Huang, P. T.; Summers, B. J.; Xu, C.; Perilla, J. R.; Malikov, V.; Naghavi, M. H.; Xiong, Y., FEZ1 Is Recruited to a Conserved Cofactor Site on Capsid to Promote HIV-1 Trafficking. *Cell Rep* **2019,** *28* (9), 2373-2385 e7.
113. Waragai, M.; Lammers, C. H.; Takeuchi, S.; Imafuku, I.; Udagawa, Y.; Kanazawa, I.; Kawabata, M.; Mouradian, M. M.; Okazawa, H., PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. *Hum Mol Genet* **1999,** *8* (6), 977-87.
114. Sun, L.; Wu, J.; Du, F.; Chen, X.; Chen, Z. J., Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **2013,** *339* (6121), 786-91.
115. Li, X. D.; Wu, J.; Gao, D.; Wang, H.; Sun, L.; Chen, Z. J., Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* **2013,** *341* (6152), 1390-4.
116. Li, X.; Shu, C.; Yi, G.; Chaton, C. T.; Shelton, C. L.; Diao, J.; Zuo, X.; Kao, C. C.; Herr, A. B.; Li, P., Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **2013,** *39* (6), 1019-31.
117. Du, M.; Chen, Z. J., DNA-induced liquid phase condensation of cGAS activates innate immune signaling. *Science* **2018,** *361* (6403), 704-709.
118. Yoh, S. M.; Schneider, M.; Seifried, J.; Soonthornvacharin, S.; Akleh, R. E.; Olivieri, K. C.; De Jesus, P. D.; Ruan, C.; de Castro, E.; Ruiz, P. A.; Germanaud, D.; des Portes, V.; Garcia-Sastre, A.; Konig, R.; Chanda, S. K., PQBP1 Is a Proximal Sensor of the cGAS-Dependent Innate Response to HIV-1. *Cell* **2015,** *161* (6), 1293-1305.
119. Edde, B.; Rossier, J.; Le Caer, J. P.; Desbruyeres, E.; Gros, F.; Denoulet, P., Posttranslational glutamylation of alpha-tubulin. *Science* **1990,** *247* (4938), 83-5.
120. Xia, P.; Ye, B.; Wang, S.; Zhu, X.; Du, Y.; Xiong, Z.; Tian, Y.; Fan, Z., Glutamylation of the DNA sensor cGAS regulates its binding and synthase activity in antiviral immunity. *Nat Immunol* **2016,** *17* (4), 369-78.
121. Sirajuddin, M.; Rice, L. M.; Vale, R. D., Regulation of microtubule motors by tubulin isotypes and post-translational modifications. *Nat Cell Biol* **2014,** *16* (4), 335-44.
122. Yoh, S.; Mamede, J.; Lau, D.; Ahn, N.; Sánchez-Aparicio, M.; Temple, J.; Fuchs, N.; Tuckwell, A.; Cianci, G.; Riva, L.; Curry, H.; Yin, X.; Gambut, S.; Simons, L.; Hultquist, J.; König, R.; Xiong, Y.; García-Sastre, A.; Böcking, T.; Hope, T.; Chanda, S. K., Recognition of HIV-1 Capsid Licenses Innate Immune Response to Viral Infection. *In press* **2021**.
123. Mizuguchi, M.; Obita, T.; Serita, T.; Kojima, R.; Nabeshima, Y.; Okazawa, H., Mutations in the PQBP1 gene prevent its interaction with the spliceosomal protein U5-15 kD. *Nat Commun* **2014,** *5*, 3822.

124. Rogowski, K.; van Dijk, J.; Magiera, M. M.; Bosc, C.; Deloulme, J. C.; Bosson, A.; Peris, L.; Gold, N. D.; Lacroix, B.; Bosch Grau, M.; Bec, N.; Larroque, C.; Desagher, S.; Holzer, M.; Andrieux, A.; Moutin, M. J.; Janke, C., A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell* **2010,** *143* (4), 564-78.
125. D'Imprima, E.; Floris, D.; Joppe, M.; Sanchez, R.; Grininger, M.; Kuhlbrandt, W., Protein denaturation at the air-water interface and how to prevent it. *Elife* **2019,** *8*.
126. Noble, A. J.; Wei, H.; Dandey, V. P.; Zhang, Z.; Tan, Y. Z.; Potter, C. S.; Carragher, B., Reducing effects of particle adsorption to the air-water interface in cryo-EM. *Nat Methods* **2018,** *15* (10), 793-795.
127. Naydenova, K.; Russo, C. J., Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat Commun* **2017,** *8* (1), 629.
128. Noble, A. J.; Dandey, V. P.; Wei, H.; Brasch, J.; Chase, J.; Acharya, P.; Tan, Y. Z.; Zhang, Z.; Kim, L. Y.; Scapin, G.; Rapp, M.; Eng, E. T.; Rice, W. J.; Cheng, A.; Negro, C. J.; Shapiro, L.; Kwong, P. D.; Jeruzalmi, D.; des Georges, A.; Potter, C. S.; Carragher, B., Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **2018,** *7*.
129. Glaeser, R. M., Proteins, Interfaces, and Cryo-Em Grids. *Curr Opin Colloid Interface Sci* **2018,** *34*, 1-8.
130. Glaeser, R. M.; Han, B. G., Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys Rep* **2017,** *3* (1), 1-7.
131. Weissenberger, G.; Henderikx, R. J. M.; Peters, P. J., Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods* **2021,** *18* (5), 463-471.
132. Chen, J.; Noble, A. J.; Kang, J. Y.; Darst, S. A., Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J Struct Biol X* **2019,** *1*.
133. Li, B.; Zhu, D.; Shi, H.; Zhang, X., Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. **2021**.
134. Popot, J. L., Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu Rev Biochem* **2010,** *79*, 737-75.
135. Palovcak, E.; Wang, F.; Zheng, S. Q.; Yu, Z.; Li, S.; Betegon, M.; Bulkley, D.; Agard, D. A.; Cheng, Y., A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *J Struct Biol* **2018,** *204* (1), 80-84.
136. Fan, X.; Wang, J.; Zhang, X.; Yang, Z.; Zhang, J. C.; Zhao, L. Y.; Peng, H. L.; Lei, J. L.; Wang, H. W., Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nature Communications* **2019,** *10*.
137. Han, Y.; Fan, X.; Wang, H.; Zhao, F.; Tully, C. G.; Kong, J.; Yao, N.; Yan, N., High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proc Natl Acad Sci U S A* **2020,** *117* (2), 1009-1014.

138. Llaguno, M. C.; Xu, H.; Shi, L.; Huang, N.; Zhang, H.; Liu, Q.; Jiang, Q. X., Chemically functionalized carbon films for single molecule imaging. *J Struct Biol* **2014,** *185* (3), 405-17.
139. Naydenova, K.; Peet, M. J.; Russo, C. J., Multifunctional graphene supports for electron cryomicroscopy. *Proc Natl Acad Sci U S A* **2019,** *116* (24), 11718-11724.
140. Rothemund, P. W., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297-302.
141. Martin, T. G.; Bharat, T. A.; Joerger, A. C.; Bai, X. C.; Praetorius, F.; Fersht, A. R.; Dietz, H.; Scheres, S. H., Design of a molecular support for cryo-EM structure determination. *Proc Natl Acad Sci U S A* **2016,** *113* (47), E7456-E7463.
142. Loayza, D.; Parsons, H.; Donigian, J.; Hoke, K.; de Lange, T., DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *J Biol Chem* **2004,** *279* (13), 13241-8.
143. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* **2012,** *109* (12), E690-7.
144. Lim, K. H.; Huang, H.; Pralle, A.; Park, S., Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnol Bioeng* **2013,** *110* (1), 57-67.
145. Chivers, C. E.; Crozat, E.; Chu, C.; Moy, V. T.; Sherratt, D. J.; Howarth, M., A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat Methods* **2010,** *7* (5), 391-3.
146. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K., A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **2003,** *21* (1), 86-9.
147. Cole, N. B., Site-specific protein labeling with SNAP-tags. *Curr Protoc Protein Sci* **2013,** *73*, 30 1 1-30 1 16.
148. DeLeeuw, L. W.; Monsen, R. C.; Petrauskas, V.; Gray, R. D.; Baranauskiene, L.; Matulis, D.; Trent, J. O.; Chaires, J. B., POT1 stability and binding measured by fluorescence thermal shift assays. *PLoS One* **2021,** *16* (3), e0245675.
149. Bromley, K. M.; Morris, R. J.; Hobley, L.; Brandani, G.; Gillespie, R. M.; McCluskey, M.; Zachariae, U.; Marenduzzo, D.; Stanley-Wall, N. R.; MacPhee, C. E., Interfacial self-assembly of a bacterial hydrophobin. *Proc Natl Acad Sci U S A* **2015,** *112* (17), 5419-24.
150. Summers, B. J.; Digianantonio, K. M.; Smaga, S. S.; Huang, P. T.; Zhou, K.; Gerber, E. E.; Wang, W.; Xiong, Y., Modular HIV-1 Capsid Assemblies Reveal Diverse Host-Capsid Recognition Mechanisms. *Cell Host Microbe* **2019,** *26* (2), 203-216 e6.