**URACILATED HIV-1 DNA FOLLOWS DIVERSE**

**FATES DURING INFECTION OF MYELOID LINEAGE CELLS**

by
Erik Christian Hansen

A dissertation submitted to Johns Hopkins University in conformity with the
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May 2016

**Abstract**


The uracil nucleobase plays a central role in the innate immune response against HIV-1
infection in resting immune cells when it is found in DNA rather than RNA. The most
well characterized uracil-centric innate immune response involves host DNA cytidine
deaminase enzymes (APOBECs), which selectively deaminate cytosine residues during
HIV-1 first strand DNA synthesis thereby rendering the viral genome nonfunctional by
G→A hypermutation. Previous work from our group suggested the presence of another
uracil-dependent HIV-1 restriction pathway that does not involve APOBEC enzymes.
The function of this pathway, which has been historically controversial, involves
incorporation of dUTP into viral DNA by reverse transcriptase to produce non-mutagenic
U/A base pairs (uracilation), which preserve the coding potential of native T/A pairs.
Notably, U/A pairs are "invisible" to normal DNA sequencing methods and their
presence, persistence and ultimate fate in proviral DNA of HIV-1 infected cells is an
important aspect of viral infection that has been largely unexplored. The limiting dNTP
pool levels, high dUTP levels and unique DNA repair capacities of non-dividing cells
such as monocytes, macrophages and dendritic cells provides a unique and poorly
understood metabolic environment for HIV-1 replication and infection.

This work describes new methods for detecting uracil bases in HIV-1 DNA and
provides a detailed examination of their fate over the course of infection in monocyte-
derived macrophage (MDM) target cells of HIV-1. We report that a major subpopulation
of MDMs has a metabolic phenotype leading to high levels of dUTP incorporation (U/A
pairs) into HIV-1 DNA during reverse transcription. Importantly, *ex vivo* analysis showed

U/A base pairs are present in short-lived blood monocytes and alveolar macrophages (but not CD4+ T cells) from antiretroviral therapy (ART)-suppressed HIV-1 infected individuals suggesting that they arise from recent passage of monocytes through a drug resistant viral reservoir. This newly defined interplay between host dNTP pools, DNA repair machinery and HIV-1 is important for understanding the potential of tissue macrophages in the establishment and maintenance of a long-term HIV reservoir.

Thesis Advisor: Professor James T. Stivers, PhD

Thesis Reader: Assistant Professor Jungsan Sohn, PhD

## Acknowledgements

I would like to thank Jim for his wonderful mentorship throughout my graduate work. Bob and Janet  Siliciano for their generosity and support.

To my familyfor all their love.

# Table of Contents

# List of Tables

## List of Figures

# Chapter 1: Introduction

**Sources of uracil in DNA**

The nucleobase uracil is a natural base of RNA but is often found in DNA through two parallel pathways that include cytosine deamination (spontaneous or enzyme catalyzed) or misincorporation of deoxyuridine 5'-triphosphate (dUTP) during DNA replication. Uracil constitutes the most common lesion in genomic DNA[1] and if not repaired, deaminated cytosine residues are inherently mutagenic where the initial U/G mispair results in C/G à T/A transition mutations upon DNA replication [2]. In contrast to cytosine deamination, the generation of U/A base pairs is not inherently mutagenic but the presence of uracil may be sensed as a cytotoxic lesion or a mutagenic event when chromosomal abasic (AP) sites are generated by the removal of uracils. Cellular organisms have evolved several universal DNA repair mechanisms to sense and remove these lesions to preserve the integrity of their genome. However, the introduction of uracil in viral genomic DNA intermediates during synthesis has become an important component of the host innate immune response against invading viruses [3]. Consequently, viruses have developed several important strategies to prevent the incorporation and persistence of uracil bases in their genome.

**Purposeful introduction of uracil APOBEC-mediated cytidine deamination**

The deamination of cytosine to uracil in DNA can arise spontaneously or enzymatically by APOBEC cytidine deaminases, which catalyze the irreversible hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine,

respectfully[3]. The apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) enzyme family is capable of deaminating cytosines within DNA or RNA. Some members of this cytidine deaminase family, including activation-induced cytidine deaminase or AID are expressed exclusively in activated B lymphocytes and are a key factor for antibody diversification[4]. AID targets single stranded DNA (ssDNA) of variable and switch regions of the host-immunoglobulin (Ig) gene locus, promoting somatic hypermutation and class-switch recombination processes that are critical for generating antibody diversity in response to antigens[5].

Perhaps the most notable of the APOBEC3 sub-family with respect to HIV-1 infection is APOBEC3G (A3G), which is a potent host restriction factor that impedes HIV-1 replication[6]. Subsequently, it was discovered that the human genome encodes seven APOBEC3 proteins: APOBEC3A, A3B, A3C, A3DE, A3F, A3G and A3H [7]. With the exception of A3H, all other members have been shown to exhibit some level of antiviral activity towards numerous retroviruses including HIV, SIV, MLV and HTLV. The antiviral activity of these APOBEC3 proteins operates through a cytidine deaminase 'editing' mechanism of viral cDNA leading to lethal G à A hypermutation in a dinucleotide C**C** or T**C** sequence specific context, where the cytosine base in bold is deaminated[8]. Aside from their antiviral effects, APOBEC3 proteins posses several alternative functions including inhibitory effects on intracellular targets MusD, IAP, Ty1, LINE-1 and an ability to prevent propagation of mobile elements in their host genomes[9].

To counteract cytidine deaminase editing and protect its genome from uracilation, HIV-1 has evolved to encode the viral infectivity factor protein, Vif [10,11]. Vif specifically targets APOBEC3G for proteasomal degradation, thus depleting APOBEC3G at the site

of virion budding and subsequent incorporation into progeny virus[6]. Importantly, all lentiviruses except EIAV encode a Vif protein, where the interaction between APOBEC3G and vif has co-evolved to be species specific[11]. Consistent with this assertion, APOBEC3 proteins from various species can be packaged *in vitro* by HIV-1, but only human APOBEC3G can be degraded by HIV-1 Vif.

**U/A base pairs arise from dUTP misincorporation**

The common RNA base uracil (U) may be substituted for thymine (T) in DNA because dUTP readily pairs with adenine (A) during DNA replication resulting in U/A base pairs. Uracil-adenine base pairs may also arise from utilization of dUTP by retroviral reverse transcriptases (RTs) that do not readily discriminate between dUTP and TTP during DNA synthesis[12]. While the initial uracil incorporation event is non-mutagenic (U/A pairs retain their coding potential), they may become mutagenic and lead to genomic instability and fragmentation when acted upon by the universal cellular DNA base excision repair machinery (BER)[13].

Incorporation of dUTP during DNA replication has been estimated to result in ~10,000 uracil residues per genome per day and represents the major source of uracil in DNA. In eukaryotes, dUTP is synthesized from the phosphorylation of dUDP, arising from UDP generated by ribonucleotide reductase (RNR) or from the phosphorylation of dUMP, which is an important intermediate in the *de novo* nucleotide biosynthesis of TTP[14]. Importantly, this constitutes a permanent source of cellular dUTP. Given that DNA polymerases from eukaryotes and viruses are not able to successfully discriminate from dUTP and TTP, the intracellular concentration and composition of deoxynucleotide

triphosphates (dNTPs) dictates the level and frequency of dUTP incorporation. Accordingly, all dividing cells express deoxyuridine triphosphate hydrolase (dUTPase) that serves to keep dUTP pool levels very small[15,16].

**Unbalanced dNTP pool levels in myeloid lineage cells**

The steady-state composition and concentration of dNTP pool levels in mammalian cells is tightly regulated due to the highly mutagenic effects of nucleotide pool imbalances in dividing cells[17] and antiviral restriction[18,19]. Under physiological conditions where dUTPase is highly expressed, the cellular concentration of dUTP and TTP are reported to be around 0.2 and 37 µM, respectively, thus the dUTP/TTP ratio is about 1/100 in dividing cells in order to prevent uracil incorporation during chromosomal DNA replication[20].

In non-dividing cells, such as terminally differentiated macrophages or other myeloid lineage cells, cellular dNTP concentrations are 6-133 fold lower than a typical dividing cell and the dUTP/TTP ratio is about 60:1 [21]. These cells maintain their low and unbalanced intracellular nucleotide pool levels through high expression levels of SAMHD1 dNTPase and limited expression of the deoxyuridine 5'-triphosphate nucleotide hydrolase (dUTPase). Consequently, viruses such as HIV-1 that are able to replicate in such an adverse dNTP-depleted environment have a high probability of incorporating dUTP during viral replication. Not surprising, viruses have evolved ways of evading extensive DNA uracilation by either preventing its introduction or facilitating its removal[22].

**SAMHD1 mediated dNTP pool depletion in quiescent immune cells**

SAMHD1 dNTPase is a recently identified viral restriction factor that is highly expressed in resting immune cells[18]. Restriction of HSV-1 and HIV-1 in quiescent immune cells has been directly linked to the expression of SAMHD1 [23]. Aside form its role as a viral restriction factor, inherited mutations in SAMHD1 lead to the severe early-onset chronic neuroinflammatory disease, Aicardi-Goutieres disease (AGS)[24]. Several alternative or secondary enzymatic activities of SAMHD1 have been described, including single-stranded (ss) DNA and RNA binding properties[25,26].

The primary and most well characterized antiviral restriction activity of SAMHD1 is the non-specific dNTP triphosphohydrolase activity that reduces all cellular dNTPs, including dUTP, to nanomolar levels which starve viral reverse transcriptase machinery of its substrates (Fig. 1) [27]. In the studies of this thesis, we defined the complex ordered-essential dNTPase activation mechanism of SAMHD1 involving two classes of activator sites and the catalytic site[28]. Briefly, occupation of the (A1) activator site by any guanine containing nucleotide (d)GTP led to dimer formation, while binding of any dNTP at the second activator (A2) and catalytic sites generated the catalytically competent tetramer. The unique allosteric regulation mechanism of SAMHD1 requires the coordinated binding of 12 nucleotides to generate the long-lived activated tetrameric species [28]. Once formed, the tetramer shows profound hysteresis and can persist for many hours even after activating nucleotides have been removed from solution. Presumably, this enzymatic property of SAMHD1 allows it to deplete nucleotides down to a level that would be impossible to achieve if activating nucleotide binding followed a simple reversible

equilibrium. We have demonstrated that the dNTP pool depletion brought about by SAMHD1 is a key factor in determinint the dUTP/dTTP ratio in quiescent immune cells.

**Cellular and viral strategies to counteract incorporated uracil**

Viruses have devised many strategies to prevent the emergence of uracilated viral genomes and eliminate uracils that have already been incorporated in to genomic DNA. Viruses that utilize a DNA intermediate during replication, such as poxvirus, herpesvirus and retroviruses encode their own viral dUTPase and/or uracil DNA glycosylase (UNG) to compensate for the low expression levels of these cellular enzymes in some immune cells[29]. An important exception is HIV-1 that successfully replicates in non-dividing immune cells but does not encode dUTPase or UNG.

As mentioned earlier, the depleted nucleotide metabolic environment of myeloid lineage cells leads to preferential utilization of dUTP instead of TTP across from adenine during reverse transcription. The dUTP pool levels are elevated due to the weak expression of dUTPase in myeloid cells [30,31]. This enzyme reduces intracellular dUTP by converting it to dUMP and inorganic pyrophosphate. Reductive methylation of dUMP by thymidylate synthase (TS) generates dTMP from dUMP, which is then phosphorylated to TTP. Thus, dUTPase reduces the relative dUTP/TTP ratio and provides dUMP for the *de novo* pathway for TTP biosynthesis. Through alternative splicing, the human *dUTPase* gene encodes a mitochondrial and nuclear localized isoform [32]. Expression of the nuclear isoform is cell cycle regulated where the highest levels of expression is seen at S phase, while nearly undetectable levels are observed in differentiated and non-dividing cells, including macrophages[33].

The second mechanism, which is responsible for removal of uracil once it is incorporated into DNA, is uracil base excision repair (UBER). UBER is the primary mechanism used to remove uracil from DNA, and is initiated by the enzyme uracil DNA glycosylase (UNG). UNG activity produces an abasic (AP) site product, which is further processed by apurinic/apyrimidinic endonuclease 1 or 2 (APE1, APE2) [34]. The resultant free 3'-OH and 5'-deoxyribose phosphate (dRP) are acted upon by the lyase activity of DNA polyermase β (pol β) for the short patch repair pathway. Alternatively, a short oligonucleotide is cleaved by the flap endonuclease 1 (FEN1) in the long patch repair pathway. Finally, the resulting gap can be filled in by polyermases pol β, δ or ε and ligated by ligase I or III[35]. DNA glycosylases involved in uracil removal are highly conserved in mammals, bacteria, yeast and the viruses mentioned previously.

Like dUTPase, the *ung* gene codes for both a mitochondrial (UNG1) and nuclear (UNG2) localized isoform through alternative splicing. UNG2 has been reported by many groups to be the DNA glycosylase solely responsible for the removal of misincorporated dUTP in the form or U/A base pairs [36]. Nuclear hUNG2 is differentially expressed through the cell cycle, it peaks at S phase in dividing cells and shows limited expression levels in non-dividing cells [37]. Studies performed in mice and humans demonstrate that a UNG deficiency leads to accumulation of genomic uracils, confirming its role in removal of uracil in DNA[38].

An analysis of primate lentiviruses (including HIV-1) demonstrated that they do not encode dUTPase or UNG-like enzymes yet are still capable of productively infecting macrophages and other non-dividing cells. The ability of these quiescent immune cells to support HIV-1 replication has led to several reports, although controversial, that suggest

8

viral packaging of host UNG2 into nascent virions [39]. The hypothesis was first supported

by a yeast two hybrid study that discovered viral protein Vpr bound to the C-terminus of

UNG[40]. This interaction appeared to be dependent on Vpr's ability to interact with the

WxxF motif contained in UNG2[41]. Although Vpr was initially suspected to play an

important role in packaging of hUNG2, many groups have since reported that viral

packaging of hUNG2 is independent of Vpr [42]. While it has been shown that Vpr

participates in reverse transcriptase fidelity, suggesting it may target hUNG2 to the viral

replication complex, the role of a Vpr-hUNG2 complex remains to be conclusively

discerned. Future experiments to further understand the function of a potential Vpr-UNG

complex need to address several important parameters before a definitive conclusion can

be made. These parameters include the use of a relevant physiological cell system,

quantitative measures of hUNG2, dNTP pool levels, and the protection of uracilated HIV

DNA from uracil excision in the absence of virally packaged hUNG2.


**General aspects of HIV-1 infection**

Human immunodeficiency virus type 1 (HIV-1) is responsible for the acquired

immunodeficiency syndrome (AIDS) that has infected close to 40 million individuals

worldwide, primarily in sub-Saharan Africa, and has killed nearly half of them [43]. HIV-1

infection of CD4+ T cells, tissue macrophages, and dendritic cells in the absence of

antiretroviral therapy (ART) leads almost undoubtedly to AIDS and death by

opportunistic infections. The advent and introduction of ART has dramatically subverted

this deadly scenario, however, the penetration of antiretroviral agents in both lymphoid

and non-lymphoid tissues is not always optimal, which leads to incomplete suppression of viral replication and possible selection of drug-resistant viral variants[44].

HIV-1 infects cells through interaction of its gp120 envelope (Env) protein with the primary entry receptor CD4 molecule and one of two chemokine coreceptors (CCR5 in the initial phase of infection, or CXCR4 in late-stage disease)[45] dualtropic R5X4 viruses are also known to exist[46] Consequently, these entry receptors restrict the spectrum of target cells that HIV-1 can infect to CD4+ T lymphocytes (whose depletion in patients is a hallmark of AIDS) and to mononuclear phagocytes (monocytes, macrophages, microglial and myeloid dendritic cells)[47].

Once in the target cell, both viral and cellular proteins form a pre-integration complex (PIC) that shuttles the viral genome to the host cell nucleus. The PIC interacts with components of the nuclear pore while reverse transcription of the viral RNA into DNA is occurring. In the nucleus, viral integrase promotes the integration of linear forms of proviral DNA in to host chromosomes[48]. At the proviral stage, HIV-1 will remain stably associated within the host cell, whereas the expression of viral RNA and proteins will depend on the interplay of its own regulatory proteins and host transcription factors. The newly synthesized structural viral proteins along with many accessory proteins will assemble at the plasma membrane along with full-length viral genomic RNA to produce new progeny virions.

The HIV life cycle has presented a rich source of targets for antiretroviral agents, which are typically delivered in combination during antiretroviral therapy (ART) [49]. ART has significantly prolonged the lives of many patients by effectively controlling viral replication over the entire lifetime of the patient [50]. However, ART has not been

successful in purging or preventing the formation of a reservoir of infected cells that become established early during primary HIV-1 infection. Although infrequent, and relatively low in numbers, these cells carry integrated, replication-competent proviruses capable of reinitiating viral reemergence and disease progression if therapy is interrupted[51]. In addition to this obstacle, differentiated tissue macrophages have the unusual feature of being able to bud out virus into the extracellular milieu like CD4+ T cells, but also in subcellular compartments recently defined as virus-containing compartments (VCCs) [52,53]. This functional feature leads to the possibility that macrophages could harbor long-lived reservoir of infectious virus that could be released upon cell death or physiological stimulation.

**HIV-1 persistence in the monocyte-macrophage lineage**

Multiple mechanisms are likely to contribute to HIV persistence during long-term antiretroviral therapy. A vast majority of CD4+ T cells that are productively infected are likely to die from virus-induced cytopathic effects, but a small subset of long-lived resting memory T cells that harbor integrated HIV DNA known to persist indefinitely, which is generally referred to as latency [54]. Memory CD4+ T cells are the largest and best characterized cellular reservoir in humans. The presence of latent proviral HIV-1 DNA in this cell population has been unambiguously proven [55]. Genetic studies have shown that during viral rebound after interrupted ART the virus could be detected from an additional distinct reservoir[56]. One possibility is that latent HIV infection may also occur in cell populations other than memory T cells, including cells from the monocyte-macrophage lineage (Fig. 1.2).

Cells of the myeloid lineage may also play an important role in controlling and propagating the initial infection and therefore contribute to the general establishment of viral reservoirs, because these cells are involved in a wide range of both innate and adaptive immune functions. In support of a contribution from myeloid lineage cells in viral persistence after ART, infected monocytes have been recovered from the blood of HIV-1 infected individuals, even from those on ART with undetectable viral loads [57-59]. Many studies have shown that monocytes harbor proviral HIV-1 DNA [60-62] Ellery et al. suggested that a minor monocyte sub-population (CD16[++]) is more permissive to infection than the more abundant CD14 [++]CD16[-] monocyte subset[63]. In ART naïve patients, HIV-1 proviral DNA is detected at very low levels (between 0.01-1%) in circulating monocytes, while CD4+ T cells likely contribute over 99% of viral particles. In this situation, viruses produced by blood monocytes may be too minor to recognize. However, in patients on suppressive ART where HIV replication in CD4+ T cells is blocked, viruses released from monocytes may become a dominant factor.

A physiologically relevant role for monocytes in HIV infection is supported by the observation that higher levels of HIV-1 transcripts and sequence evolution is observed in CD14+ monocytes than in resting CD4+ T cells of patients on ART [60]. This finding suggests a higher level of HIV-1 replication in monocytes compared to resting CD4+ T cells. Importantly, HIV-infected circulating blood monocytes may serve as an indirect source of HIV-1 by carrying virus in the circulation and then differentiating into tissue macrophages where HIV may replicate for long periods due to the resistance of macrophages to viral induced apoptosis (Trojan horse hypothesis)[64]. Given their short half-life in peripheral blood (1-3 days) prior to differentiating in to macrophages [65], the

finding of persistent HIV-1 proviruses in blood monocytes suggests they were recently infected by passage through a tissue reservoir. Thus, ongoing renewal of an infected monocyte/macrophage pool may be important for seeding HIV-1 into sanctuary sites in various tissues.

Macrophages are myeloid lineage cells of the innate immune system that can be found in virtually every tissue of the body and are responsible for myriad of functions, from tissue homeostasis and repair to recognition and destruction of invading pathogens and tumor cells [50,66]. Macrophages originate from either primitive embryonic precursors or from infiltrating bone stem cell-derived monocytes, leading to monocyte-derived macrophages (MDMs)[67]. *In vitro* derived MDMs are the most common surrogate model of tissue macrophage physiopathology, including HIV-1 infection[67]. The function and phenotype of MDMs differ widely between and within tissues, which has important implications in normal physiology and disease such as HIV-1. HIV-1 infection of macrophages was first described in the 1980s from *ex vivo* analysis of tissues from post mortem AIDS patients[68]. Originally a high frequency of infected macrophages was observed in brain, lung and secondary lymphoid tissue [69]. However, recent studies have shown infected macrophages in many tissues at all stages of disease, which persist under ART[70,71]. The *in vivo* morphological features of infected macrophages of AIDS patients are recapitulated by *in vitro* infection of human MDMs with macrophage-tropic viral strains suggesting that MDMs are a reasonable model system for understanding *in vivo* HIV infection of macrophages [72].

Given the difficulty in accessing relevant tissues, the role that macrophages and other myeloid cells play in shaping and maintaining the viral reservoir has been difficult

to discern. Accordingly, research efforts have focused on accessible tissue reservoirs where most of the virus is thought to reside, particularly where there are dense collections of target cells that might support cell-to-cell transfer of virus[73] Specifically, understanding the role of macrophages as a viral reservoir and the way in which signaling between myeloid cells and CD4+ T cells leads to HIV persistence have become critically important.

**Macrophages and viral pathogenesis**

Despite the successes of ART, standard therapies do not fully restore health or a normal immune status in HIV infected individuals and patients still experience co-morbidities, including increased cognitive impairment, cardiovascular disease, and bone disorders[74]. In addition to their role in the spread and persistence of HIV-1, infection of macrophages may be directly involved in promoting disease through the activation of inflammatory processes. Consistent with this assertion, HIV-1 infection of the central nervous system (CNS) is widely associated with inflammation leading to pathology ranging from encephalitis to mild but progressive congestive dysfunction, associated with neuronal death[75]. Although ART generally reduces the burden of CNS disease, low levels of residual inflammation resulting from persistent infection of resident macrophages may continue to cause progressive disease.

Both respiratory and bowel dysfunction have been associated with HIV-1 infection. Impaired macrophage function has been implicated in lower respiratory tract infections as evidenced by the fact that HIV-1 infected alveolar macrophages have reduced phagocytic activity both in ART-suppressed and naïve donors [76]. HIV-1 infection

of the gut-associated lymphoid tissue (GALT) results in damage to the immunological and physical barriers, which can lead to microbial translocation promoting chronic immune activation [77]. Given that monocytes are recruited by inflammation to the site of tissue damage and also polarize these incoming cells according to the local environment, this would lead to amplification of ongoing inflammatory processes in the GALT.

Any local or global inflammation associated with HIV-1 infection is a serious risk factor for cardiovascular disease (CVD) and macrophages are proposed to play a central role in CVD pathogenesis. Human trials have revealed correlations between duration of HIV-1 infection, markers of inflammation, the presence of inflammatory macrophages in the vasculature and CVD[57]. A recent study with implications for both reservoir formation and pathogenesis of osteolytic bone disease is the infection of osteoclasts in both ART-suppressed and naïve donors. Osteoclasts are myeloid lineage cells that can be derived *in vitro* from blood monocytes, express CD4 and CCR5 and are readily infected by HIV-1 [78]. These infected osteoclasts upregulate markers of activation and increased bone resorption activity, suggesting a direct role in bone disease.


**Barriers to viral eradication and a cure for HIV-1**

Despite the advent and effectiveness of ART, the ability for HIV-1 to establish latent infection remains the key obstacle to viral eradication and a cure for HIV-1. If ART therapy is discontinued, even after years of viral suppression, latently infected cellular reservoirs become reactivated leading to viral rebound. These cellular reservoirs are dormant, long-lived cells that harbor replication-competent viruses that evade immune detection because viral proteins are expressed very infrequently or not at all [79]. Most of

the attention in this regard has focused on CD4+ T cell reservoirs in which the virus can reside in a latent state that will be difficult to eliminate. While the majority of research has focused on CD4+ T cells as the primary cellular reservoir of HIV, there is evidence to support other sites as potential cellular or anatomical reservoirs, including (GALT), peripheral blood, bone marrow, and the brain.

Macrophages are infected *in vivo* in a tissue-specific manner. GALT has long been associated with HIV-1 infection and is abundant in macrophages [80]. *Ex vivo* analysis of human GALT macrophages showed that they are relatively refractory to infection compared to MDMs, and this phenotype was supported by exposure of MDMs to TGFβ, which down regulated CD4 and CCR5 expression and inhibited viral infection [81]. Despite the comparative refractory nature of GALT macrophages to infection, HIV-1 has been detected in macrophages from human duodenum tissues from ART-suppressed donors [80]. Further, macrophages that were proximal to the rectum exhibited increased CCR5 expression and also enhanced HIV-1 susceptibility compared to colon-resident cells, reinforcing the importance of the tissue-specific macrophage phenotype in infection. Given that CCR5 directly mediates macrophages susceptibility to HIV-1, its increased expression through the clinical course of infection suggests a greater role for macrophages in advanced stages of disease[82,83].

The brain has been recognized as an organ target for HIV-1 macrophage infection since the 1980's. The principal target cells in the brain are the microglia and perivascular macrophages, which are infected rapidly after transmission, although the source of infecting virus is unclear [84,85]. While access to the human CNS during acute infection is prohibitively challenging due to the blood-brain barrier, it seems likely that once infected,

the CNS will remain so for the lifetime of the host, implying that this is an important viral reservoir[75]. Consistent with this assertion, recent work has shown that genetically distinct HIV-1 variants exist in the CNS and plasma, and importantly, viral RNA can be detected in the CSF of ART-suppressed donors when viral RNA is undetectable in the plasma [56].

Aside from the CNS, the majority of the limited amount of information on HIV-1 infection of human tissue macrophages has been derived from few anatomic sites that are accessible with minimally invasive procedures. On accessible site is lung alveolar macrophages, obtained from bronchioalveolar lavage (BAL) [86]. These are targets of HIV-1 with important implications for local tissue damage and viral reservoir formation. Fluorescence in situ hybridization (FISH) revealed macrophages transcribing HIV-1 in small alveolar macrophages[76]. A study involving ART-suppressed donors, showed that while most of the patients analyzed had undetectable plasma viral loads, (16/23) had detectable proviral HIV-1 DNA in alveolar macrophages, of which eight had measureable viral RNA[77]. Again, compartmentalization of HIV infection within the lung of ART-treated donors is strong evidence that this tissue may well be an important persistent viral reservoir.

As the HIV research field investigates strategies to eliminate cellular reservoirs that persist, even during aggressive ART therapy, the nature of the cellular and tissue reservoirs that sustain viral persistence becomes an important question. Central to this question and the role that resident macrophages may play in viral eradication is the significance of their compartmentalization in drug-resistant tissues. While ART has been extremely successful in preventing disease progression, primarily through preservation of

the CD4+ T cell population, many of these drugs have low penetrance and are not maintained at high levels in many tissue sub-compartments harboring macrophages[87].

Given that macrophages are long-lived and can support ongoing viral replication in drug-resistant tissues warrants additional attention. The persistence of intracellular HIV-1 in sanctuary sites such as the CNS, where the penetration of some antiviral drugs is restricted by the blood-brain barrier, is considered a relevant obstacle to the treatment-induced elimination of infection. Several studies have demonstrated that viral protease and reverse transcriptase inhibitors are less effective in chronically infected macrophages compared to CD4+ T cells[21]. This reinforces the hypothesis that mononuclear phagocytes as well as microglia could accumulate infectious virus, which would remain invisible to immune recognition, while virus production at these sites could be less affected by systemic administration of antiretroviral agents.

**A hidden HIV-1 restriction and persistence mechanism in macrophages**

Previous reports on the role of viral DNA uracilation, host uracil base excision repair and HIV-1 infection of resting immune cells have been inconsistent and controversial [reviewed in [50]]. In the studies described here, we evaluated the mechanism by which uracils are introduced in viral DNA of MDMs and explored the biological consequences. We now report the basic elements of a restriction/persistence pathway involving incorporation of dUTP into HIV DNA and their excision by the host UBER pathway of infected MDMs (Fig. 1.3). Surprisingly, this pathway only operates in a sub-set of myeloid lineage cells and not in activated or resting CD4+ T cells.

We have found that monocyte-derived macrophages (MDMs) consist of two sub-populations of cells distinguished by their ability to express GFP upon infection with a GFP-containing reporter virus. The GFP $^-$ sub-population contains high dUTP levels and uracilated viral DNA products after infection, whereas viral DNA products in the GFP $^+$ sub-population do not contain.

The fundamental biochemical characteristic of the GFP $^-$ sub-population relevant to uracilation is their elevated dUTP/TTP level. This phenotype is brought about by their high expression levels of SAMHD1 dNTPase and by their low or undetectable dUTPase activity. Although GFP $^+$ macrophages and CD4+ T cells also share high expression levels of SAMHD1, they both exhibit higher expression of dUTPase, which depletes the dUTP pool and precludes the uracilation phenotype.

**Impact of UBER on virus reemergence and sequence diversification**

We estimate, based on comparing the measured levels of early and proviral DNA levels in the GFP$^-$ MDMs, that ~99% of the viral DNA is destroyed before the integration step. The viral DNA products that escape initial restriction can faithfully integrate into the host cell DNA with similar efficiency as non-uracilated DNA. Uracilated proviruses can then persist as U/A pairs, undergo faithful repair by UBER to yield normal T/A base pairs, or suffer error-prone repair resulting in viral mutagnesis or abortive infection. RNA pol II, not reverse transcriptase, appears to be the primary source of viral mutagenesis in infected MDMs. The mutational frequency can be augmented by the pro-inflammatory (M1) cytokine, IFNγ, while the anti-inflammatory (M2) IL-2 appears to significantly reduce viral mutagenesis. The combination of transcription-coupled repair and RNA

polymerase II associated errors may play an important role in viral evolution in macrophages independent of reverse transcription.

**Use uracil to elucidate target cellular reservoirs of HIV-1**

We have detected U/A base pairs in short-lived blood monocytes from ART-suppressed HIV infected individuals suggesting that they arise from recent passage of monocytes through a drug resistant viral reservoir. This work may help define a more precise description of where the virus exists during effective therapy and how these reservoirs acquire HIV. Interestingly, on top of the genetic aspects of U/A pairs is their potential to down regulate gene expression through epigenetic silencing [88], which could allow transcriptionally silenced proviruses to endure and participate in the establishment and persistence of infection. Uracilation of viral DNA may define a new mechanism that maintains viral quiescence in persistently infected cells. Knowledge about pathways leading to uracil incorporation and persistence may provide needed insights into how latency is established and maintained, and might lead to the development of new therapeutics that target myeloid cell populations that are enriched for latent HIV.

**Figures**
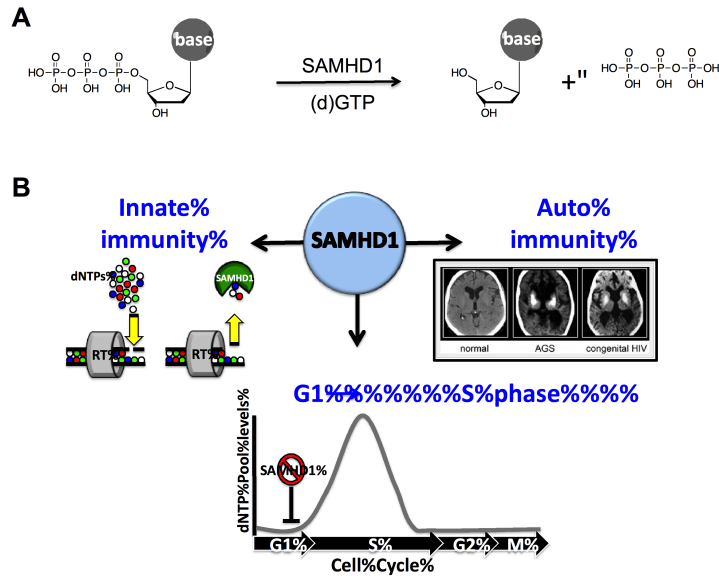


**Figure 1.1: Diverse roles for SAMHD1.** (A) Non-specific dNTP triphosphohydrolase activity of SAMHD1 reduces all dNTPs to the deoxynucleoside and inorganic tripolyphosphate. (B) Involvement of SAMHD1 in innate and adaptive immune responses and cell cycle regulation.
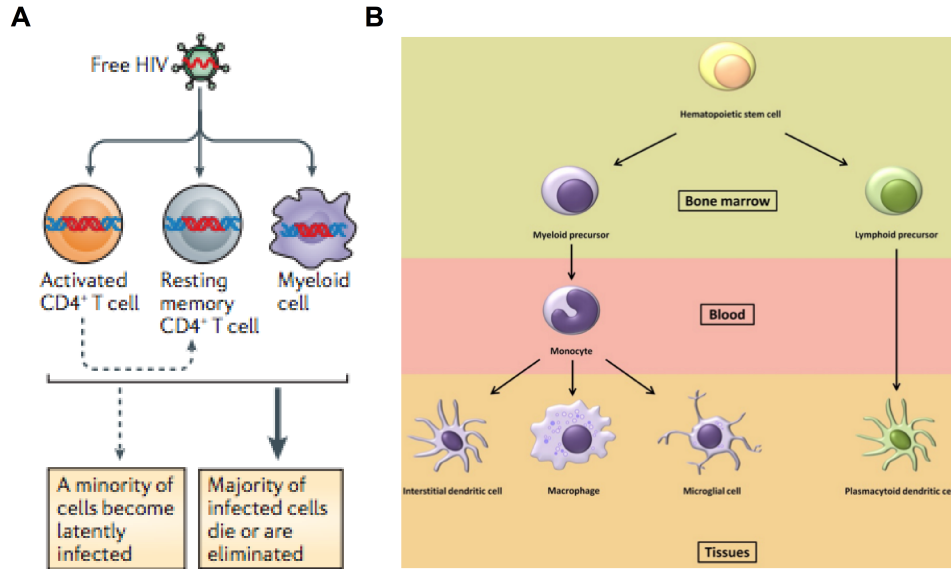
**Figure 1.2: Cellular targets of HIV-1** (A). Latent HIV infection can be established in T cell and myeloid cell reservoirs. All cells from the monocyte-macrophage lineage appear to derive from a same progenitor multipotent cell, the hematopoietic stem cell (HSC). (B). The HSC, located in the bone marrow, may differentiate either into a myeloid or a lymphoid precursor, setting up the divergence between the myeloid (blue) and plasmacytoid (green) lineage. The myeloid precursor is then able to migrate into the blood stream and differentiate into a monocyte. Depending to the location, the monocytes become either interstitial dendritic cells, macrophages or microglial cells. (Adapted from Deeks et al.[89] and Le Douce et al.[90])
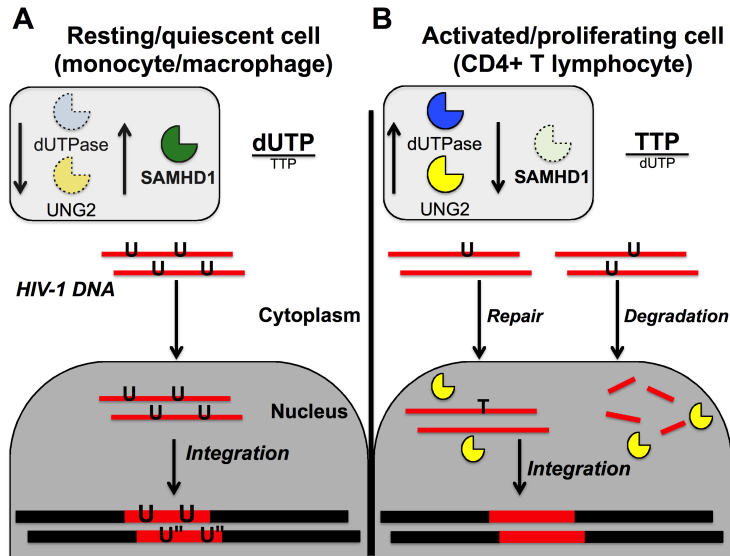
**Figure 1.3: Uracil incorporation/persistence pathway in monocyte-macrophage lineage cells.** (A). The unique nucleotide metabolic environment (low expression of dUTPase and high SAMHD1) in monocytes and macrophages leads to the preferential incorporation of dUTP in viral cDNA. Modest UNG2 expression levels in these cells promote the survival and persistence of uracilated viral DNA. (B) Depletion of dUTP through high levels of dUTPase is the major barrier excluding dUTP incorporation. Robust levels of nuclear localized UNG2 results in extensive damage and fragmentation of uracil containing viral DNA.

。

◦ **References**

1. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
2. Andersen, S. Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis* **26,** 547–555 (2004).
3. Sousa, M. M. L., Krokan, H. E. & Slupphaug, G. DNA-uracil and human pathology. *Mol Aspects Med* **28,** 276–306 (2007).
4. Muramatsu, M. *et al.* Specific Expression of Activation-induced Cytidine Deaminase (AID), a Novel Member of the RNA-editing Deaminase Family in Germinal Center B Cells. *J Biol Chem* **274,** 18470–18476 (1999).
5. Bhagwat, A. S. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair* **3,** 85–89 (2004).
6. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **9,** 1404–1407 (2003).
7. Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat Rev Immunol* **4,** 868–877 (2004).
8. Harris, R. S. *et al.* DNA Deamination Mediates Innate Immunity to Retroviral Infection. *Cell* **113,** 803–809 (2003).
9. Esnault, C. C. *et al.* APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* **433,** 430–433 (2005).
10. Marin, M., Rose, K. M., Kozak, S. L. & Kabat, D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nature Med* **9,** 1398–1403 (2003).
11. Mariani, R. *et al.* Species-Specific Exclusion of APOBEC3G from HIV-1 Virions by Vif. *Cell* **114,** 21–31 (2003).
12. Greagg, M. A. *et al.* A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc Natl Acad Sci USA* **96,** 9045–9050 (1999).
13. Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J. & Haynes, R. H. dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *The EMBO Journal* **12,** 4425–4431 (1993).
14. Mathews, C. K. DNA precursor metabolism and genomic stability. *The FASEB Journal* **20,** 1300–1314 (2006).
15. el-Hajj, H. H., Zhang, H. & Weiss, B. Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *J. Bacteriol.* **170,** 1069–1075 (1988).
16. Vértessy, B. G. & Tóth, J. Keeping Uracil Out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases. *Acc Chem Res* **42,** 97–106 (2009).
17. 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 USA* **110,** 12942–12947 (2013).

18. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480,** 379–382 (2011).

19. Baldauf, H.-M. *et al.* SAMHD1 restricts HIV-1 infection in resting CD4+ T cells. *Nat Med* **18,** 1682–1689 (2012).

20. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem* **140,** 1–22 (1994).

21. Gavegnano, C. & Schinazi, R. F. Antiretroviral therapy in macrophages: implication for HIV eradication. *Antivir Chem Chemother* **20,** 63–78 (2009).

22. Priet, S., Sire, J. & Querat, G. Uracils as a Cellular Weapon Against Viruses and Mechanisms of Viral Escape. *Current HIV Research* **4,** 31–42 (2006).

23. Kim, E. T., White, T. E., Brandariz-Nunez, A., Diaz-Griffero, F. & Weitzman, M. D. SAMHD1 restricts herpes simplex virus type 1 (HSV-1) in macrophages by limiting DNA replication. *J Virol* **87,** 12949–12956 (2013).

24. Powell, R. D., Holland, P. J., Hollis, T. & Perrino, F. W. Aicardi-Goutieres Syndrome Gene and HIV-1 Restriction Factor SAMHD1 Is a dGTP-regulated Deoxynucleotide Triphosphohydrolase. *J Biol Chem* **286,** 43596–43600 (2011).

25. Beloglazova, N. *et al.* Nuclease Activity of the Human SAMHD1 Protein Implicated in the Aicardi-Goutieres Syndrome and HIV-1 Restriction. *J Biol Chem* **288,** 8101–8110 (2013).

26. 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* **43,** 6486–6499 (2015).

27. Amie, S. M. *et al.* Anti-HIV Host Factor SAMHD1 Regulates Viral Sensitivity to Nucleoside Reverse Transcriptase Inhibitors via Modulation of Cellular dNTP Levels. *J Biol Chem* (2013). doi:10.1074/jbc.M113.472159

28. 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 USA* **111,** E1843–E1851 (2014).

29. Fleming, S. B., Lyttle, D. J., Sullivan, J. T., Mercer, A. A. & Robinson, A. J. Genomic analysis of a transposition-deletion variant of orf virus reveals a 3.3 kbp region of non-essential DNA. *J Gen Virol* **76 ( Pt 12),** 2969–2978 (1995).

30. Lichtenstein, D. L. *et al.* Replication in vitro and in vivo of an equine infectious anemia virus mutant deficient in dUTPase activity. *J Virol* **69,** 2881–2888 (1995).

31. Steagall, W. K., Robek, M. D., Perry, S. T., Fuller, F. J. & Payne, S. L. Incorporation of uracil into viral DNA correlates with reduced replication of EIAV in macrophages. *Virology* **210,** 302–313 (1995).

32. Ladner, R. D. & Caradonna, S. J. The Human dUTPase Gene Encodes both Nuclear and Mitochondrial Isoforms: Differential expression of the isoforms and characterization of a cDNA encoding the mitochondrail species. *J Biol Chem* **272,** 19072–19080 (1997).

33. Ladner, R. D. *et al.* dUTP nucleotidohydrolase isoform expression in normal and neoplastic tissues: association with survival and response to 5-fluorouracil in colorectal cancer. *Cancer Res* **60,** 3493–3503 (2000).

34. Almeida, K. H. & Sobol, R. W. A unified view of base excision repair: Lesion-

dependent protein complexes regulated by post-translational modification. *DNA Repair* **6,** 695–711 (2007).

35. Sire, J., Querat, G., Esnault, C. & Priet, S. Uracil within DNA: an actor of antiviral immunity. *Retrovirology* **5,** 45–12 (2008).

36. Grogan, B. C., Parker, J. B., Guminski, A. F. & Stivers, J. T. Effect of the Thymidylate Synthase Inhibitors on dUTP and TTP Pool Levels and the Activities of DNA Repair Glycosylases on Uracil and 5-Fluorouracil in DNA. *Biochemistry* **50,** 618–627 (2011).

37. Hagen, L. *et al.* Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. *The EMBO Journal* **27,** 51–61 (2008).

38. Imai, K. *et al.* Human uracil–DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* **4,** 1023–1028 (2003).

39. Willetts, K. E. *et al.* DNA repair enzyme uracil DNA glycosylase is specifically incorporated into human immunodeficiency virus type 1 viral particles through a Vpr-independent mechanism. *J Virol* **73,** 1682–1688 (1999).

40. Bouhamdan, M. *et al.* Human immunodeficiency virus type 1 Vpr protein binds to the uracil DNA glycosylase DNA repair enzyme. *J Virol* **70,** 697–704 (1996).

41. Begum, N. A. *et al.* Requirement of non-canonical activity of uracil DNA glycosylase for class switch recombination. *J Biol Chem* **282,** 731–742 (2007).

42. Kaiser, S. M. & Emerman, M. Uracil DNA Glycosylase Is Dispensable for Human Immunodeficiency Virus Type 1 Replication and Does Not Contribute to the Antiviral Effects of the Cytidine Deaminase Apobec3G. *J Virol* **80,** 875–882 (2005).

43. Prejean, J. *et al.* Estimated HIV Incidence in the United States, 2006–2009. *PLoS ONE* **6,** e17502–13 (2011).

44. Robillard, K. R. *et al.* Role of P-Glycoprotein in the Distribution of the HIV Protease Inhibitor Atazanavir in the Brain and Male Genital Tract. *Antimicrob Agents Chemother* **58,** 1713–1722 (2014).

45. Wu, Y. & Yoder, A. Chemokine Coreceptor Signaling in HIV-1 Infection and Pathogenesis. *PLoS Pathog* **5,** 1–8 (2009).

46. Ghezzi, S. *et al.* Inhibition of R5X4 Dualtropic HIV-1 Primary Isolates by Single Chemokine Co-receptor Ligands. *Virology* **280,** 253–261 (2001).

47. Pantaleo, G., Graziosi, C. & Fauci, A. S. New concepts in the immunopathogenesis of human immunodeficiency virus infection. *N Engl J Med* **328,** 327–335 (1993).

48. Carter, C. A. & Ehrlich, L. S. Cell Biology of HIV-1 Infection of Macrophages. *Annu Rev Microbiol* **62,** 425–443 (2008).

49. Arts, E. J. & Hazuda, D. J. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb Perspect Med* **2,** a007161–a007161 (2012).

50. Sattentau, Q. J. & Stevenson, M. Macrophages and HIV-1: An Unhealthy Constellation. *Cell Host Microbe* **19,** 304–310 (2016).

51. Siliciano, R. F. & Greene, W. C. HIV Latency. *Cold Spring Harb Perspect Med* **1,** a007096–a007096 (2011).

52. Deneka, M., Pelchen-Matthews, A., Byland, R., Ruiz-Mateos, E. & Marsh, M. In macrophages, HIV-1 assembles into an intracellular plasma membrane domain

containing the tetraspanins CD81, CD9, and CD53. *J Cell Biol* **177,** 329–341 (2007).

53. Jouve, M., Sol-Foulon, N., Watson, S., Schwartz, O. & Benaroch, P. HIV-1 Buds and Accumulates in 'Nonacidic' Endosomes of Macrophages. *Cell Host Microbe* **2,** 85–95 (2007).

54. Hermankova, M. *et al.* Analysis of Human Immunodeficiency Virus Type 1 Gene Expression in Latently Infected Resting CD4+ T Lymphocytes In Vivo. *J Virol* **77,** 7383–7392 (2003).

55. Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387,** 183–188 (1997).

56. Dahl, V. *et al.* An Example of Genetically Distinct HIV Type 1 Variants in Cerebrospinal Fluid and Plasma During Suppressive Therapy. *J Infect Dis* **209,** 1618–1622 (2014).

57. Crowe, S. The contribution of monocyte infection and trafficking to viral persistence, and maintenance of the viral reservoir in HIV infection. *J Leukoc Biol* **74,** 635–641 (2003).

58. Wang, T. *et al.* Successful Isolation of Infectious and High Titer Human Monocyte-Derived HIV-1 from Two Subjects with Discontinued Therapy. *PLoS ONE* **8,** e65071–12 (2013).

59. Llewellyn, N. *et al.* Continued evolution of HIV-1 circulating in blood monocytes with antiretroviral therapy: genetic analysis of HIV-1 in monocytes and CD4+ T cells of patients with discontinued therapy. *J Leukoc Biol* **80,** 1118–1126 (2006).

60. Zhu, T. *et al.* Evidence for Human Immunodeficiency Virus Type 1 Replication In Vivo in CD14+ Monocytes and Its Potential Role as a Source of Virus in Patients on Highly Active Antiretroviral Therapy. *J Virol* **76,** 707–716 (2002).

61. Crowe, S. M. & Sonza, S. HIV-1 can be recovered from a variety of cells including peripheral blood monocytes of patients receiving highly active antiretroviral therapy: a further obstacle to eradication. *J Leukoc Biol* **68,** 345–350 (2000).

62. Lewin, S. R. *et al.* HIV-1 DNA and mRNA concentrations are similar in peripheral blood monocytes and alveolar macrophages in HIV-1-infected individuals. *AIDS* **12,** 719–727 (1998).

63. Ellery, P. J. *et al.* The CD16+ Monocyte Subset Is More Permissive to Infection and Preferentially Harbors HIV-1 In Vivo. *J Immunol* **178,** 6581–6589 (2007).

64. González-Scarano, F. & Martín-García, J. The neuropathogenesis of AIDS. *Nat Rev Immunol* **5,** 69–81 (2005).

65. Jakubzick, C. *et al.* Minimal Differentiation of Classical Monocytes as They Survey Steady-State Tissues and Transport Antigen to Lymph Nodes. *Immunity* **39,** 599–610 (2013).

66. Martin, C. J., Peters, K. N. & Behar, S. M. ScienceDirectMacrophages clean up: efferocytosis and microbial control. *Curr Opin Microbiol* **17,** 17–23 (2014).

67. Haldar, M. & Murphy, K. M. Origin, development, and homeostasis of tissue-resident macrophages. *Immunol Rev* **262,** 25–35 (2014).

68. Perno, C. F. *et al.* Therapeutic strategies towards HIV-1 infection in macrophages. *Antiviral Res* **71,** 293–300 (2006).

69. Cory, T. J., Schacker, T. W., Stevenson, M. & Fletcher, C. V. Overcoming

pharmacologic sanctuaries. *Curr Opin HIV AIDS* **8,** 190–195 (2013).

70. Gavegnano, C., Kennedy, E. M., Kim, B. & Schinazi, R. F. The Impact of Macrophage Nucleotide Pools on HIV-1 Reverse Transcription, Viral Replication, and the Development of Novel Antiviral Agents. *Mol Biol Int* **2012,** 1–8 (2012).

71. Gerngross, L. & Fischer, T. Evidence for cFMS signaling in HIV production by brain macrophages and microglia. *J Neurovirol* **21,** 249–256 (2014).

72. Arrildt, K. T. *et al.* Phenotypic Correlates of HIV-1 Macrophage Tropism. *J Virol* **89,** 11294–11311 (2015).

73. Ballweber, L. *et al.* Vaginal langerhans cells nonproductively transporting HIV-1 mediate infection of T cells. *J Virol* **85,** 13443–13447 (2011).

74. Guaraldi, G. *et al.* Premature Age-Related Comorbidities Among HIV-Infected Persons Compared With the General Population. *Clin Infect Dis* **53,** 1120–1126 (2011).

75. Fois, A. F. & Brew, B. J. The Potential of the CNS as a Reservoir for HIV-1 Infection: Implications for HIV Eradication. *Curr Opin HIV AIDS Rep* **12,** 299–303 (2015).

76. Jambo, K. C. *et al.* Small alveolar macrophages are infected preferentially by HIV and exhibit impaired phagocytic function. *Mucosal Immunol* **7,** 1116–1126 (2014).

77. Cribbs, S. K., Lennox, J., Caliendo, A. M., Brown, L. A. & Guidot, D. M. Healthy HIV-1-Infected Individuals on Highly Active Antiretroviral Therapy Harbor HIV-1 in Their Alveolar Macrophages. *AIDS Res Hum Retroviruses* **31,** 64–70 (2015).

78. Gohda, J. *et al.* HIV-1 replicates in human osteoclasts and enhances their differentiation in vitro. *Retrovirology* **12,** 12 (2015).

79. Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nature Medicine* 1–9 (2009).

80. Zalar, A. *et al.* Macrophage HIV-1 infection in duodenal tissue of patients on long term HAART. *Antiviral Res* **87,** 269–271 (2010).

81. Shen, R. *et al.* Stromal Down-Regulation of Macrophage CD4/CCR5 Expression and NF-κB Activation Mediates HIV-1 Non-Permissiveness in Intestinal Macrophages. *PLoS Pathog* **7,** e1002060–11 (2011).

82. Tuttle, D. L., Harrison, J. K., Anders, C., Sleasman, J. W. & Goodenow, M. M. Expression of CCR5 increases during monocyte differentiation and directly mediates macrophage susceptibility to infection by human immunodeficiency virus type 1. *J Virol* **72,** 4962–4969 (1998).

83. de Roda Husman, A. M., Blaak, H., Brouwer, M. & Schuitemaker, H. CC chemokine receptor 5 cell-surface expression in relation to CC chemokine receptor 5 genotype and the clinical course of HIV-1 infection. *J Immunol* **163,** 4597–4603 (1999).

84. Yukl, S. A. *et al.* A comparison of methods for measuring rectal HIV levels suggests that HIV DNA resides in cells other than CD4+ T cells, including myeloid cells. *AIDS* **28,** 439–442 (2014).

85. Yukl, S. A. *et al.* The Distribution of HIV DNA and RNA in Cell Subsets Differs in Gut and Blood of HIV-Positive Patients on ART: Implications for Viral Persistence. *J Infect Dis* **208,** 1212–1220 (2013).

86. Costiniuk, C. T. & Jenabian, M.-A. The lungs as anatomical reservoirs of HIV infection. *Rev Med Virol* **24,** 35–54 (2013).

87.    Gavegnano, C. *et al.* Cellular Pharmacology and Potency of HIV-1 Nucleoside Analogs in Primary Human Macrophages. *Antimicrob Agents Chemother* **57,** 1262–1269 (2013).

88.    Weil, A. F. *et al.* Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc Natl Acad Sci USA* **110,** E448–E457 (2013).

89.    Deeks, S. G. Towards an HIV cure: a global scientific strategy. *Nat Rev Immunol* 1–8 (2012).

90.    Le Douce, V., Herbein, G., Rohr, O. & Schwartz, C. Molecular mechanisms of HIV-1 persistence in the monocyte-macrophage lineage. *Retrovirology* **7,** 32 (2010).

**Chapter 2: GTP activator and dNTP substrates of HIV-1 restriction factor**

**SAMHD1 generate a long-lived activated state**

*This chapter has been reproduced, with modification, from* Hansen E.C., Seamon K.J., Cravens S.L., Stivers J.T. Proc. Natl. Acad. Sci. U.S.A. 2014;111:E1843-E1851. *(http://www.pnas.org/content/111/18/E1843.full)*

◦

◦ **Introduction**

The steady-state composition and concentration of deoxynucleotide triphosphate pools in mammalian cells are highly regulated because of the mutagenic consequences of dNTP imbalances in dividing cells (1, 2) as well as the important antiviral effects of dNTP pool depletion in quiescent cells (3, 4). In all cell types, the ultimate pool balance is determined by dNTP-dependent regulatory pathways that affect the activities of enzymes involved in both synthesis and degradation of dNTPs (5–7). The most important highly up-regulated synthetic enzyme during S phase of dividing cells is the R1/R2 isoform of ribonucleotide triphosphate reductase, which ensures that dNTP precursors are plentiful for DNA synthesis (8). However, in quiescent cells of the immune system (resting CD4+ T cells, macrophages, and dendritic cells), where dNTP pools are ~10-fold lower than dividing cells, the ultimate pool levels are likely determined by a balance between the activities of the R1/p53R2 isoform of ribonucleotide triphosphate reductase and the degradative dNTP triphosphohydrolase sterile α-motif/histidineaspartate domain-containing protein 1 (SAMHD1) (9). The highly dynamic nature of dNTP pools demands finely tuned mechanisms for feedback regulation of these enzymes by dNTPs as well as coarse regulatory mechanisms (posttranslational modifications, transcriptional regulation, and proteasomal targeting) that serve to turn these activities on and off at appropriate

stages of the cell cycle and in specific cell types (10, 11). dNTP triphosphohydrolase enzymes, such as SAMHD1, are conserved from bacteria to humans and carry out the unusual conversion of dNTPs to the nucleoside (dN) and tripolyphosphate (12, 13). The tripolyphosphate is likely degraded to 3Pi through the action of plentiful cellular pyrophosphatases or tripolyphosphatases (14). Presumably, degradation to the level of nucleoside rather than deoxynucleoside diphosphate (dNDP) or deoxynucleoside monophosphate (dNMP) is to make the process energetically or kinetically costly to reverse, with the additional possibility that the neutral nucleoside will be irreversibly transported out of the cell (9). The unusual triphosphohydrolase activity of SAMHD1 in quiescent immune cells has received significant attention, because HIV-1 and HSV-1 are severely restricted in their ability to infect quiescent cells that have severely depressed dNTP pools, which has been directly linked to SAMHD1 enzymatic activity (15, 16). Mutations in the SAMHD1 gene have also been linked to Aicardi–Goutières syndrome, a rare genetic autoimmune encephalopathy with a chronic inflammatory pathology that resembles congenital viral infections (17). As would be expected, resting CD4+ T cells from Aicardi–Goutières syndrome patients are permissive to HIV-1 infection (18). A key question with SAMHD1 is how its activity is tuned in response to dNTP pool levels. Early reports established that the dNTPase activity of SAMHD1 required the presence of dGTP as an activator when dATP, dTTP, or dCTP was used as a substrate (15, 17). Subsequently, it was reported that the dGTP activator/ substrate can be replaced by GTP, which serves as an activator, but not a substrate (19). Two recent structures of SAMHD1 in complex with dNTPs show a tetrameric quaternary structure with dGTP molecules bound to the $(A1A2)_4$ activator sites and one dNTP bound to each of four catalytic sites

31

(20). Comparison with an earlier dimeric structure of the free enzyme suggests that activator and/or substrate binding drive the enzyme into the tetrameric form (15, 21). The importance of this enzyme in innate immunity (15), autoimmunity (17), and control of the transition from G1 to S phase in the cell cycle (5) makes an understanding of its regulatory and catalytic mechanism of interest. In addition, such mechanistic work can also provide useful insights to guide the design of mechanism-based activators and inhibitors of SAMHD1 that could have research or therapeutic uses. In this study, we have characterized the coordinated effects of activator and substrate binding on oligomerization and activation of the enzyme and discovered a long-lived activated state of SAMHD1 that is not in equilibrium with free activator nucleotides. We propose that GTP combined with any dNTP serve as the preferred activators for SAMHD1 in the cell and that the long-lived activated state is key for efficient dNTP depletion.

**Materials and Methods**

**SAMHD1 Enzyme Cloning and Overexpression.** SAMHD1 full length (1881 bp) construct was PCR amplified using the forward primer (Full length; 5'-GTAACTCATATGCAGCGAGCCGATTCC-3', and the reverse primer 5'-GCACCTGGATCCCTACATTGGGTCATCTTT-3') from the cloned full length cDNA vector from the Invitrogen Ultimate ORF collection (clone IOH55544). SAMHD1 gene was ligated into a pET19b-His$_{10}$-PPS (Novagen) plasmid (Nde1 and BamH1) and the sequence was confirmed by sequencing both DNA strands. Full length construct was expressed as a N-terminal His$_{10}$-PPS fusion. *Escherichia coli* BL21-DE3 cells (Novagen) were transformed with the pET19b-His$_{10}$-PPS plasmid and grown in LB medium at 37°C.

Once an $OD_{600} = 0.5$ was achieved the temperature was reduced to 16°C on ice. Expression was induced by addition of 0.25 mM IPTG. After expression for 24 h at 16°C, cells were harvested via centrifugation (4,000 x g) and frozen at -80°C. Cells were resuspended in lysis buffer containing [50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 10% glycerol] and one protease inhibitor cocktail tablet (Sigma). Cells were lysed by addition of 0.5 mg/mL hen egg white lysozyme. The resulting cell lysate was clarified via centrifugation (30,000 x g), filtered with a 0.2 μm syringe top filter and loaded onto 10 ml of Ni-NTA resin (Qiagen) that had been equilibrated in buffer A [50 mM Tris-HCl (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol] at 4°C. Unbound material was removed with 3 column volumes of buffer A supplemented with 50 mM imidazole and bound SAMHD1 protein was eluted using a linear gradient of buffer A containing 0-100% 500 mM imidazole. Fractions containing SAMHD1 were pooled and incubated with PreScission protease (PPS) for 1 h at 4°C to remove N-terminal $His_{10}$ tag. Crude protein was diluted 10-fold into buffer A [50 mM Tris-HCl (pH 7.0), 20 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol] and loaded onto a SP-sepharose cation-exchange column (GE Healthcare). Bound protein was eluted with a 0-100% linear gradient of buffer A containing 500 mM NaCl. Fractions containing protein were analyzed via SDS-PAGE and judged to be ~99% pure. Purified proteins were buffer exchanged into storage buffer [50 mM Tris-HCl (pH 7.5), 150 mM KCl, 5 mM $MgCl_2$, 1 mM DTT and 10% glycerol] by chromatography using sephacryl 200HR gel filtration resin (Sigma-Aldrich). Protein concentrations were determined by absorbance measurements at 280 nm using the calculated (Protparam tool, ExPASy) molar extinction coefficients for full length human SAMHD1 monomer ($\varepsilon = 76,500$ M$^{-1}$

33

cm$^{-1}$).  Protein yields were typically 20 mg/L of bacterial culture.  Purified proteins were stored at -80°C in small portions. Experiments were initiated by thawing a single (20 μL) aliquot, stored at -20°C, and used over the course of three days before being discarded.

**Steady-State Kinetic Measurements.** Standard reaction conditions for steady-state kinetic measurements were 0.01 to 5 mM GTP (activator), 50 mM Tris-HCl (pH 7.5), 50 mM KCl, 5 mM MgCl$_2$, and 0.5 mM TCEP in a 12 μL total reaction volume at 22°C. Concentrations of the [5-$^3$H] dUTP or [8-$^3$H] dGTP substrates were varied in the range 0.01 to 5 mM and standard reactions were initiated by the addition of SAMHD1.  Two microliter samples were removed at indicated times and quenched by spotting onto a C18-reversed phase thin layer chromatography (TLC) plate.   The TLC plate was developed in 50 mM KH$_2$PO$_4$ (pH 4.0) to separate substrate [dGTP ($R_f = 0.80$) or dUTP ($R_f = 0.97$)] from products [dG ($R_f = 0.20$), dU ($R_f = 0.54$)].  Plates were exposed to a tritium sensitive screen for 5 h and then scanned on a Typhoon phosphoimager (GE Healthcare) and the counts present in the substrate and product were quantified using the program Quantity One (Bio-Rad).  The amount of product formed at each timepoint was calculated from the ratio (cpm dN product)/(cpm dNTP substrate + cpm dN product) (intial [dNTP]).  Initial rates of product formation were obtained from the slopes of linear plots of [dN] versus time at reaction extents of less than 20%.  The reaction rates (μM/s) were plotted versus dNTP substrate concentration at various concentrations of GTP activator as described.   For GTP activation with dUTP as the substrate, the kinetic parameters were determined by fitting to an ordered essential activation mechanism (eq 1), where [S] is dUTP, [A] is GTP, $K_m^{app, dUTP}$ is the apparent Michaelis constant at a given concentration of GTP activator, $K_{act}^{app, GTP}$ is the activation constant for GTP at a

given [S], and $V_{max}^{app,\,dUTP}$ is the apparent maximal velocity for dUTP hydrolysis at a given activator concentration.

$$v = \frac{V_{max}^{app,dUTP}}{K_m^{app,dUTP}} \cdot \frac{[S]}{1+[S]} \tag{1}$$

$$V_{max}^{app,dUTP} = \frac{k \cdot k_{dUTP}}{\left(k' + \frac{k \cdot k_{dUTP}}{[S]}\right)} \tag{2}$$

$$K_i^{app,dUTP} = K_i^{dUTP}\left(1 + \frac{k^{app,dUTP}}{[S]}\right) \tag{3}$$

$$K_{dGTP}^{app,dUTP} = \frac{k \cdot k_{dGTP}}{\left(k' + \frac{[S]}{k \cdot k_{dUTP}}\right)} \tag{4}$$

**Mechanism of Activation and Inhibition by dGTP.** dGTP is both a cis-activator and a substrate and sigmoidal plots of velocity versus [dGTP] would be expected. However, sigmoidicity was not observed and the steady-state velocity data in the presence and absence of the trans-activator GTP were fitted to a simple hyperbolic expression (eq 1). To investigate the activation and inhibition of dUTP hydrolysis by dGTP, we used a fixed concentration of the dUTP substrate (1 mM) and varying concentrations dGTP in the range 5 μM to 5 mM. These data were fitted to the mechanism in eqs 5-8 using numerical integration methods and nonlinear least squares optimization to the data (30).

$$E + dGTP \xrightleftharpoons{K_{act}} EdGTP \tag{5}$$

$$EdGTP + dUTP \xrightleftharpoons{K_m} EdGTPdUTP \tag{6}$$

$$EdGTP + dGTP \xrightleftharpoons{K_i} EdGTPdGTP \tag{7}$$

$$EdGTPdUTP \xrightarrow{k_{cat}} EdGTP + dU \tag{8}$$
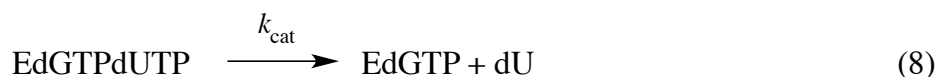
**dGTPαS: Inhibition, Activation and Reactivity.** Inhibition by dGTPαS was evaluated under reaction conditions containing SAMHD1 (0.5 μM), fixed dUTP (1 mM) substrate while varying the concentrations of dGTPαS (5 μM – 5 mM). Since dGTP behaves both as a cis-activator and a substrate, we presumed that dGTPαS would behave similarly. Therefore, the data were fitted by simulation to the mechanism in eqs 5-8 to obtain $K_{act}^{dGTPαS}$ and $K_i^{dGTPαS}$.

**Dilution-Jump Kinetic Measurements.** In the standard assay SAMHD1 (10 μM) was incubated for 10 seconds with varying concentrations of GTP, dUTP, and/or dGTP prior to a 100-fold dilution into standard reaction buffer containing [5$^3$H] dUTP (1 mM). Time points were then quenched on a TLC plate and product formation was quantified as described above. Plots of [dU] against time were fitted to the equation [dU] = $A$[1 - exp(-$k_{inact}t$)] + $vt$, where $A$ is the burst amplitude (μM), $k_{inact}$ is the rate constant for the burst decay (s$^{-1}$), and $v$ is the linear steady-state rate (μM/s). Variations of the standard assay were performed as indicated in the text.

**Dilution-Jump Glutaraldehyde Cross-linking.** The standard dilution-jump procedure

was performed except that samples from the diluted reactions were taken at various times and mixed with 50 mM glutaraldehyde. Samples were incubated at 22°C for 15 min, quenched by addition of 1 M Tris, pH 7.5 and loaded on to a 4-12% Bis-Tris denaturing polyacrylamide gel. Protein samples were visualized by silver staining. Briefly, gels were fixed overnight in a 50 mL solution containing 50% methanol, 10% acetic acid and 0.02% paraformaldehyde. Gels were washed three times for 20 min in 20% ethanol, followed by 5 min incubation in sensitizing solution containing 0.02% sodium thiosulfate. Sensitizing solution was removed by washing twice with 50 mL of ddOH and immediately transferred to an equivalent volume of staining solution containing 0.2% silver nitrate and 0.03% paraformaldehyde. Finally, gels were developed for 2-5 min in a 6% sodium carbonate, 0.02% paraformaldehyde (50 mL) solution. Developer was quenched by direct addition of an equivalent volume of 50% methanol and 10% acetic acid. Gels were imaged immediately using an Epson V750 flat-bed scanner. Quantitation of monomer, dimer and tetramer populations was performed using Quantity-One (Bio-rad) software.

**Binding Stoichiometry and Kinetic Release of Activator Nucleotides.** The binding of nucleotides was measured by mixing SAMHD1 (10 $\propto$M) with [$\gamma$-$^{32}$P] GTP (0.5 mM) and [5-$^3$H] dUTP (1 mM) in standard reaction buffer. This solution was immediately applied to a buffer-equilibrated Micro Bio-Spin P-30 gel filtration spin column (Bio-Rad) and the protein was eluted according to manufacturer instructions. After 2 min, the eluted protein was filtered through a second Micro Bio-Spin P-30 column. After each spin step the protein eluate was cross-linked with glutaraldehyde to quantify the percentage of tetrameric SAMHD1, as described above. The remaining solution present after the first

and second columns (50-60 ∝L) was added to 10 mL of scintillation fluid and $^{32}$P and $^{3}$H decays were counted using appropriate channel windows on a Beckman LS-6500 Scintillation Counter. Control reactions without enzyme were run to determine background from unretained nucleotides (<1%). The recovery of SAMHD1 from the column was 96%. After correcting for the background and the protein recovery, the stoichiometry of bound dUTP and GTP per monomer subunit of tetrameric SAMHD1 was calculated.

**Analytical Ultracentrifugation.** Sedimentation velocity experiments were performed on a Beckman XL-I analytical ultracentrifuge equipped with an An-Ti 60 rotor and interference optics. All runs were performed at 50,000 rpm and 20°C with scans taken every 30 seconds for 9 hours. SAMHD1 protein samples were dialyzed overnight at 22°C against a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM KCl, 5 mM MgCl$_2$, 0.5 mM DTT, and 5% glycerol. After loading the cells, the rotor was allowed to thermally equilibrate for 3 hours prior to the start of the run. SEDNTERP was used to determine partial specific volume, density, and viscosity of the protein and buffer (31). These parameters were then used in SEDFIT to fit the scan data to a c(S) distribution (32). To determine the dissociation constants for the monomer-dimer ($K_D$) and dimer-tetramer equilibria ($K_T$), scans at different concentrations were globally fit using SEDANAL by assigning an arbitrary slow dissociation rate constant of $10^{-5}$ sec$^{-1}$. The 95% confidence intervals were determined from the integrated F-statistics module in SEDANAL (33).

**Computational Modeling.** Structural models of SAMHD1 bound to dUTP/GTP and GTP/GTP were made by modification of the crystal structure containing tetrameric SAMHD1 bound to dGTP in both activator sites (PDB ID: 4MZ7). The nucleobase and

sugar moieties were built using Discovery Studio (Accelrys) and then positioned into the allosteric pockets in the same planes and configurations as the original ligands. No further refinement of the dUTP structure was required. However, the GTP/GTP ribose sugars showed steric conflict and were subjected to 1000 cycles of steepest descent energy minimization using the CHARMM27 molecular dynamics force field within HyperChem7 (HyperCube, Inc.) in an unsuccessful attempt to alleviate the steric clash. Visualization of the resultant structures was done using PyMOL.

**Results**

**Activation of dUTP Hydrolysis by GTP.** The complex dGTP/GTP-dependent activation mechanism for dNTP hydrolysis and the possibility that SAMHD1 might exist in multiple active oligomeric states presents a challenging system for designing kinetic experiments that reveal useful insights into function and mechanism. Since activation by dGTP is complicated by the fact that it is also a substrate, we began by investigating activation by GTP, which does not compete for substrate binding to the catalytic site.

The activation of dUTP hydrolysis by GTP was studied using $^3$H-labeled dUTP (**Fig. 1$A$, $B$**). dUTP was selected for these studies because it is present at high concentrations in resting immune cells that express SAMHD1 and it reacts similarly as all other dNTPs (**Fig. S1**). Variable concentrations of dUTP in the presence of five fixed GTP concentrations (0.01 – 5 mM) were incubated with SAMHD1 (0.5 µM monomer). The concentration dependence of the velocities showed no evidence of sigmoidicity and accordingly, the data were globally fit to eq 1 (**Fig. 1$C$**). The $K_m^{app, dUTP}$ values decreased

as the GTP activator was increased giving a limiting $K_m^{dUTP} = 1.5 \pm 0.1$ mM at saturating

GTP activator, with $K_{act}^{GTP} = 163 \pm 30$ µM, and $V_{max}^{dUTP}/[\text{monomers}] = k_{cat}^{dUTP} = 5 \pm 0.2$

$s^{-1}$. Normalization of $V_{max}^{dUTP}$ to monomer concentration is justified because plots of

velocity against SAMHD1 monomer concentration were linear in the

range 0.2 to 10 uM when the GTP activator concentration was $\geq 50$ µM. (**Fig. S2**).

A secondary plot of $K_m^{app}$ against 1/[GTP] activator concentration clearly shows that

the catalytic site has undetectable affinity for dUTP in the absence of GTP (**Fig. 1D**), and

that GTP acts to lower the $K_m^{app, \, dUTP}$, without changing $V_{max}^{app, \, dUTP}$ (**Fig. S3A**).

Therefore, GTP is a $K_m$-type essential activator (A), and high dUTP substrate

concentrations (S) always pull the equilibrium towards the active EAS complex even at

low activator concentrations. The cellular implication of this mechanism is that as long

as the activator concentration exceeds that of the enzyme, dNTP binding can in principle

pull the system all the way to the activated EAS state.

To confirm the ordered essential activation mechanism we also performed a kinetic

analysis where dUTP hydrolysis was measured at fixed dUTP concentrations (0.2 to 5

mM) and the GTP activator concentration was varied in the range 0.01 to 5 mM (**Fig.**

**1E**). As expected, the secondary replot of $1/K_{act}^{app, \, GTP}$ against [dUTP] was linear, with

the *y*-intercept providing the true $K_{act}^{GTP} = 133 \pm 10$ µM (**Fig. 1F**). The important

implication from this result is that GTP binds to activator site(s) of SAMHD1 monomers

or dimers in the absence of substrate. The secondary plot $1/V_{max}^{app, \, dUTP}$ against 1/[dUTP]

also showed the expected linear response from an ordered-essential activation mechanism

(**Fig. S3B**), and order-of-addition experiments demonstrated that these kinetic properties

of SAMHD1 are invariant (**Fig. S4**). We note that oligomerization is enzyme

concentration dependent, and therefore, the $k_{act}$ parameters for activators of SAMHD1 are apparent values that hold for the dilute concentrations of monomer used in these experiments.

To summarize, SAMHD1 shows classical ordered binding of activator and substrate. Despite the complex nature of SAMHD1 activation these data indicate that binding of the dUTP substrate to the four catalytic sites, as well as GTP activator binding to the guanine-specific A1 activator sites, *occur independently without any detectable cooperativity*. Accordingly, in terms of kinetic models it is sufficient to consider only a phenomenological EAS complex. We defer to the discussion the likely mechanistic basis for the apparent absence of cooperativity in this reaction. In addition, these data in isolation do not provide information as to whether the A2 activator site is occupied by GTP, or alternatively, dUTP.

**Activation and Inhibition by dGTP and dGTPαS.** Unlike GTP, dGTP can activate its own hydrolysis and is known to serve as a trans-activator for hydrolysis of other substrate dNTPs at low concentrations (17). However, it should also act as an inhibitor of dNTP hydrolysis at higher concentrations (20, 22). To explore these properties of dGTP we determined the concentration dependence of the initial rates of dGTP hydrolysis at a SAMHD1 concentration of 0.2 ∝M and variable concentrations of dGTP in the range 5 μM to 5 mM (**Fig. 2***A*, closed circles). Although a self-activating substrate is expected to show a sigmoidal concentration dependence to the initial velocities (23), these data did not require the use of the Hill equation ($K_{0.5}^{dGTP} = 1.9 \pm 0.3$ mM, and $k_{cat}^{dGTP} = 2 \pm 0.2$ s$^{-1}$). We discuss later how the unusual activation mechanism of SAMHD1 can mask expected cooperativity (see Discussion).

We found that GTP was unable to stimulate the hydrolysis of dGTP beyond the self-activation provided by dGTP alone. Addition of a saturating 5 mM fixed concentration of GTP to reactions where dGTP was the variable substrate resulted in a simple hyperbolic velocity profile that was indistinguishable from dGTP alone (**Fig. 2***A*, open circles). The steady-state kinetic parameters for dGTP hydrolysis derived from this data are very similar to those observed above in the GTP activation of the dUTP reaction and the dGTP self-activation data ($K_{0.5}^{dGTP}$ = 2.1 ± 0.4 mM, $k_{cat}^{dGTP}$ = 2 ± 0.5 s$^{-1}$). Additionally, we discovered that the steady-state turnover of SAMHD1 with a low concentration of dGTP (25 ∝M) could not be stimulated by spiking with 5 mM GTP activator (**Fig. 2***B*). This intriguing finding suggested that GTP cannot access its A1 activator site during steady-state turnover of the activated enzyme form. The basis for this intriguing property is explored in the experiments below.

We further explored whether dGTP could activate the hydrolysis of dUTP at low concentrations, and then inhibit the reaction of dUTP by competitive binding to the catalytic site as the concentration of dGTP was increased. These outcomes were born out in steady-state rate measurements (**Fig. 2***C*), where the dUTP concentration was held constant at 1 mM and dGTP was varied in the concentration range 5 μM to 5 mM. These data were simulated by fixing the $K_m^{dUTP}$ at the value obtained with the activator GTP, while allowing $K_{act}^{dGTP}$, $K_i^{dGTP}$ and $k_{cat}^{dUTP}$ to float (see eqs 5-8). If $K_m^{dUTP}$ was allowed to float, similar results were obtained but the fitting errors increased. The constants obtained for dGTP were $K_{act}^{dGTP}$ = 65 ± 33 μM, $K_i^{dGTP}$ = 146 ± 74 μM, and $k_{cat}^{dUTP}$ = 5.5 ± 1.5 s$^{-1}$. Thus, GTP and dGTP have similar activation constants for dUTP hydrolysis but dGTP results in less activation because it is also a strong competitive inhibitor ($K_i^{dGTP}$ =

1/10 $K_m^{dUTP}$).

The crystal structure of SAMHD1 bound to dGTPαS shows that the *R* stereoisomer of this analogue occupies all activator sites and catalytic sites of the tetramer suggesting that it should serve as a stable dGTP analogue for mechanistic studies (20). We found that racemic dGTPαS was a very slow substrate in the absence and presence of trans-activation by GTP ($t_{1/2}$ = 24 hrs) (**Fig. S5**), which is consistent with the crystal structure showing that dGTPαS can bind to the A1 and A2 activator sites as well as the catalytic sites (20). Supporting the suggestion that dGTPαS can self-activate, it also served as a weak trans-activator for dUTP hydrolysis at low concentration ($K_{act}^{dGTPαS}$ = 8 ± 2 µM, $k_{cat}^{dUTP}$ = 0.9 ± 0.1 s$^{-1}$) and an inhibitor at higher concentrations ($K_i^{dGTPαS}$ = 67 ± 19 µM) (**Fig. 2C**). These properties of dGTPαS, which may underestimate its potency by a factor of two due to the use of a racemic mixture, makes it useful for comparing the discrete effect of a non-substrate activator (GTP) with that of activators that are known to bind to all activator and substrate sites of SAMHD1 (dGTP, dGTPαS) (22, 20).

**Long-lived Activated States of SAMHD1.** The observation that a high concentration of GTP was not able to further stimulate SAMHD1 dGTPase activity at low dGTP concentrations (**Fig. 2B**), required us to consider that the activated form of the enzyme was not in equilibrium with free activator during turnover. To test this we performed a series of dilution-jump experiments where a high concentration of SAMHD1 (10 µM) was pre-incubated with high concentrations of GTP activator and/or dUTP substrate and then rapidly diluted 100-fold into a solution that contained millimolar concentrations of the dUTP substrate *but no activator* (**Fig. 3A**). This experiment probes the kinetic stability of activated enzyme form(s) that are generated in the pre-jump conditions.

When GTP and dUTP were present in the pre-jump solution, a rapid initial burst rate of dUTP hydrolysis was observed in the post-jump reaction ($v_i = 0.28 \pm 0.02$ µM/s) that was 6-fold greater than a standard steady-state reaction of fully activated SAMHD1 under equivalent conditions (5 mM GTP, 1 mM dUTP, 100 nM SAMHD1 monomer) (**Fig. 3*A***). This burst phase corresponds to over 2500 turnovers of enzyme active sites. We attribute the post-jump burst-decay rate constant to the slow disappearance of the highly active form of SAMHD1 generated in the pre-jump solution ($k_{inact} = 0.0011 \pm 0.0001$ s$^{-1}$; $t_{1/2} = 10 \pm 1$ min). At the conclusion of the burst-decay period, a second lower activity form persisted and produced a linear rate of dUTP hydrolysis for at least 6 hours ($v_{ss} = 0.02 \pm 0.002$ µM/s). This rate is comparable to the calculated steady-state rate of fully-activated SAMHD1 under these conditions ([GTP] = 5 mM, [dUTP] = 1 mM and [E] = 100 nM monomer), indicating that this long-lived form is the steady-state active form of SAMHD1. Thus, using pre-jump conditions containing both dUTP substrate and GTP activator, two distinct forms of SAMHD1 can be detected in the post-jump reaction. These forms differ with respect to their half-lives for decay (~ 10 min and greater than 6 hours), as well as their catalytic activities.

The post-jump results were very dependent on the composition of the pre-jump solution. When GTP alone was present in the pre-jump, the same two kinetic phases were observed as with the combination [GTP + dUTP]. However, when the pre-jump contained only SAMHD1, or enzyme and dUTP alone, the burst was abolished and exceedingly slow linear rates were observed in the post–jump reaction with dUTP (**Fig. 3*A***, closed and open squares). With only GTP in the pre-jump, the post-jump burst amplitude was reduced by 50%, but the burst-decay rate constant and steady-state rate

(normalized for active enzyme) were essentially unchanged ($k_{inact} = 0.0010 \pm 0.0003$ s$^{-1}$, $v_{ss} = 0.02 \pm 0.001$ µM/s). Consistent with the surprising finding that GTP could not stimulate steady-state turnover using low concentrations of dGTP (see **Fig. 2B**), addition of GTP to the post-jump solution containing 1 mM dUTP did not enhance the burst properties or steady-state linear rate obtained from a pre-jump solution that contained both GTP and dUTP (**Fig. 3B**). This finding establishes that the GTP activated form of SAMHD1 behaves like the dGTP activated enzyme and is not in communication with free activators during steady-state turnover.

When the substrate/self-activator dGTP was present in the pre-jump, the post-jump behavior was identical to the combination [dUTP + GTP] (**Fig. 3C**). We could establish a requirement for dGTP binding to the A2 or catalytic sites by inserting a pre-jump delay before dilution, thereby consuming any free dGTP that was not protected by being bound to the enzyme. As the pre-jump delay was increased, the post-jump burst amplitude decreased until it reached a plateau level of 50% of maximal. This is identical to that observed above when only GTP activator was present in the pre-jump (**Fig. 3A**). These results are consistent with a model where GTP and dGTP behave identically when interacting with the guanine specific A1 site, but binding of dGTP or dUTP to the A2 and catalytic sites is required to achieve full-activation.

We conclude from the above dilution-jump results that activator and substrate must be present in the pre-jump solution to fully generate the two active enzyme species that are detected in the post-jump phase. Maximal activation is provided by (i) the combination GTP (A1 site only) and dUTP (A2 and catalytic sites), or (ii) dGTP alone (all sites). Lesser activation is observed with GTP alone, which can only occupy the A1

45

site. Further evidence for this interpretation is found below and in the Discussion. Importantly, once activation occurs SAMHD1 is no longer in equilibrium with free activator while it performs steady-state turnover. We report further salient kinetic properties of the dilution-jump reactions in **Fig. S6*A-D***.

**(d)NTP Induced Changes in Oligomeric State of SAMHD1.** We hypothesized that the slowly decaying active species present in the dilution-jump experiments might reflect dissociation of the active tetramer, or alternatively, dissociation of one or more activating nucleotides. To elucidate the oligomeric species that were present during the different phases of the post-jump period we performed an analogous procedure except that the samples taken from the post-jump reactions at various times were mixed with 50 mM glutaraldehyde (**Fig. 4**). The standard dilution-jump cross-linking experiment involves pre-incubation of SAMHD1 (10 μM) with various combinations of GTP activator (0.5 mM), dUTP (1 mM), dGTP (1 mM), dGTPαS (5 mM), for 10 s before rapid 100-fold dilution into a post-jump solution containing 1 mM dUTP substrate.

   Samples subjected to this procedure and cross-linked immediately after dilution are shown in **Figure 4*A*** after separation of monomer, dimer and tetramer using denaturing polyacrylamide gel electrophoresis and visualization with silver staining. The salient findings from these data are (i) SAMHD1 in the absence of guanine nucleotides exists as an inactive monomer-dimer equilibrium even in the presence of 1 mM dUTP in the post-jump, (ii) highly efficient tetramer formation requires either dGTP alone, [GTP + dUTP], or dGTPαS in the pre-jump solution, and (iii) GTP alone in the pre-jump induces about 25% the amount of tetramer as the combined presence of [GTP + dUTP] as long as dUTP is present in the post-jump. These effects of nucleotides on the formation of tetramer

exactly parallel their effects on activity in the initial times of the post-jump kinetic assay.

We performed a series of cross-linking studies as a function of post-jump time to show that the kinetic phases in the kinetic dilution-jump experiments were not correlated with dissociation of the tetrameric state of SAMHD1. The relative amounts of each oligomeric species at each post-jump time were quantified by imaging the silver stained gels as summarized in **Fig. 4B**. In all conditions where tetramer was formed in the pre-jump step (dGTP, GTP, dGTPαS, [GTP + dUTP]), it was found to persist intact during the entire post-jump burst-decay time (60 min). *Thus, the initial burst-decay period cannot be attributed to dissociation of the tetramer.* Moreover, tetramer was still present in significant amounts six hours after the dilution jump when dUTP + GTP or dGTP were present in the pre-jump (**Fig. 4B**). Thus, the observed changes in activity must result from other more subtle alterations in the enzyme structure that are not detected by the cross-linking method.

We were curious if the highly active form of SAMHD1 present in the activator-free post jump solution retained tightly bound activator nucleotides. To address this question we performed reactions similar to the pre-jump kinetics where SAMHD1 was mixed with [γ-$^{32}$P] GTP and [5-$^{3}$H] dUTP and then subjected to a spin column gel filtration chromatography step to separate free and bound (d)NTPs (**Fig. S7A**). To evaluate the time course for release of the bound nucleotides that were retained after the first column step, a second spin column step was performed on the flow-through fraction from the first column. The simultaneous use of [γ-$^{32}$P] GTP and [5-$^{3}$H] dUTP and dual-channel scintillation counting allowed the precise ration of both nucletides to be determined. Further, determination of the amount of tetramer present after each column step allowed

47

for estimation of the stoichiometry of GTP and dUTP binding to each monomer subunit of the tetramer (see **Methods** and **Fig. S7B**). After the first column, an average of 1.3 and 0.87 equivalents of [5-$^3$H] dUTP and [ $\gamma$-$^{32}$P] GTP were bound per monomer of the tetramer form, consistent with the A1 and A2 sites being occupied by GTP and dUTP, respectively. Since the catalytic site has a high $K_m$ we assume that dUTP is rapidly released from this site during the spin column step leaving only the A2 site occupied. The time course for release of the bound GTP and dUTP was largely complete in the one-minute delay time before running the second spin column, from which we estimate a half-life for activator release of about one minute **(Fig. S7B)**. We conclude that GTP and dUTP activators are bound in one-to-one stoichiometry with each monomer subunit of the tetramer, and that their release greatly precedes that of tetramer dissociation.

We used the velocity sedimentation method to complete the characterization of the linkage between the oligomeric equilibria of SAMHD1 and binding of the nucleotides GTP and dGTPαS. The raw scan data were converted to a c($S$) distribution and then fitted to a monomer-dimer-tetramer model as described in Methods. SAMHD1 in the concentration range 2 to 8 µM total monomer largely existed in a monomer-dimer equilibrium in the absence of nucleotides (**Fig. 5A** (red) and **Fig. S8**). The monomer form dominated at the lower concentrations, but at 8 µM total monomer the monomer-dimer equilibrium slightly favored dimer ($K_D$ = 2.3 µM; 95% CI: 2.2 to 2.4). Even at the highest monomer concentration tested, less than 4% of total monomer was in the tetramer form in the absence of nucleotide ($K_T \geq$ 30 µM). Addition of 1 mM GTP pushed the monomer-dimer equilibrium significantly towards dimer ($K_D^{GTP}$ = 0.38 µM; 95% CI: 0.33 to 0.42), but induced a much smaller increase in tetramer ($K_T^{GTP}$ = 16 µM; 95% CI:

15 to 17) (**Fig. 5***A*).  In contrast, the predominant effect of 1 mM dGTPαS was to shift the dimer-tetramer equilibrium to favor tetramer ($K_T^{dGTP\alpha S}$ = 0.41 μM; 95% CI: 0.38 to 0.43, $K_D^{dGTP\alpha S}$ = 0.89 μM; 95% CI: 0.77 to 1.0) (**Fig. 5***C*).  These effects of GTP and dGTPαS on the oligomeric equilibria recapitulate their effects on the dilution-jump kinetics and cross-linking and support the model where GTP induces dimerization, but occupation of the A2 and substrate sites is required for tetramerization.

∘     **Discussion**

**Nucleotide-Dependent Ordered Assembly of SAMHD1.** These diverse measurements allow for the construction of a minimal model for nucleotide dependent activation of SAMHD1 (**Fig. 6**).  In the absence of nucleotides, SAMHD1 is trapped in an equilibrium between monomer (M) and dimer (D) forms and is unable to generate active tetramer (T) (**Fig. 6***A*).  Binding of GTP to the A1 site (but not the A2 or the catalytic sites, see below) strongly shifts the equilibrium towards dimer, with little effect on the tetramerization equilibrium (**Fig. 6***A*).  The structural basis for the increased dimer affinity apparently results from docking of the guanine base of GTP into the highly specific guanine binding site on one monomer and the formation of electrostatic interactions between its triphosphate group and the second monomer of the dimer (**Fig. S9***A, B*) (22).  Hence, GTP serves as a nucleotide tether to bring together monomers.  Since GTP concentrations exceed that of dGTP by about 1000-fold in the cell (3), and there is no large difference in the activation constants for these nucleotides, the implication of these findings is that GTP occupies the guanine-specific A1 site under physiological conditions, *not dGTP*.  An additional advantage of using GTP as the primary activator is that GTP levels are not

49

depleted by SAMHD1, allowing the activated tetramer to persist (see below).

Based on the limited ability of GTP alone to induce tetramerization and the results from structural modeling (**Fig. S9C**), we propose that the 2-hydroxyl group of GTP sterically prevents binding of GTP to both the A1 and A2 activator sites. Thus, as suggested above, GTP is restricted to the A1 activator site, and the A2 site must be occupied by a dNTP co-activator to efficiently produce tetramer (i.e. dUTP in the studies here). The contention that any dNTP can bind to the A2 site is supported by a crystal structure of SAMHD1 with dGTP bound to the specific A1 site and dATP bound to the A2 site (22), structural modeling showing that the A2 site nicely accommodates dUTP (**Fig. S9D, E**), the binding of one equivalent of dUTP to each monomer only in the presence of GTP (**Fig. S7B**), and the dilution-jump cross-linking showing that when dUTP is added in addition to the A1 site activator GTP, tetramer is efficiently formed (**Fig. 4**).

To understand the combined effect of activator and substrate on oligomerization we used the relatively inert dGTPαS analogue (20). Binding of dGTPαS shifted the equilibrium dramatically towards tetramer as compared to GTP alone, an effect that resulted from the same increase in dimer affinity observed with GTP, and a further enhancement of the tetramerization constant (**Fig. 5, 6**). This result is consistent with the crystal structure showing that dGTPαS can occupy the A1, A2 and catalytic sites, and that binding of dGTPαS to the A2 site forms bridging interactions within the tetramer (dimer-dimer) interface involving its base and triphosphate groups (**Fig. S10A**) (20).

**Ordered Pathway for Activation.** The combined activation and oligomerization data using GTP alone, [GTP + dUTP], dGTP and dGTPαS define an ordered pathway for

nucleotide activation of SAMHD1.  The first event is occupation of the A1 site by GTP, which promotes dimerization, followed by occupancy of the A2 site by a co-activator dNTP, which is thermodynamically coupled with binding of substrate dNTP to the catalytic site.  The combined effect of dNTP binding to the A2 and catalytic sites pulls the entire set of equilibria towards active tetramer.  This is the structural and temporal basis for the ordered-essential activation mechanism.  The dilution jump and direct binding measurements of GTP and dUTP indicate that once activating nucleotides are bound they are not required to maintain tetramer stability.  This result leads to the unanticipated conclusion that activating nucleotides serve much like chaperones in promoting tetramer formation, but once tetramer is formed it persists without the requirement for bound activator nucleotides.

Quite surprisingly, none of the experimental data for activation and catalysis revealed any evidence for cooperativity.  This non-intuitive and perplexing finding was brought into focus using kinetic simulations that included the step of forming a long-lived activated state such as that depicted in **Fig. 6B**.  These simulations revealed that *cooperativity would only be apparent in the pre-steady state regime*  when such a mechanism operates.  Since initial velocities are assumed to be instantaneous, the hysteresis in the system would not be observed and simple steady-state initial velocity equations would still hold.

**Long-lived SAMHD1.**  An ordered-essential activation mechanism involving GTP activation and a long-lived tetramer has several important implications for the cellular activity of SAMHD1. Since all dNTPs are substrates for SAMHD1, the driving force for formation of the active Michaelis complex would be provided by the concentration of the

51

entire dNTP pool. The observation that an active tetrameric form of SAMHD1 is stable for over six hours in the absence of activating nucleotides is relevant to the condition where dNTP pools have been largely depleted by SAMHD1. If SAMHD1 immediately dissociated to inactive dimer and monomer states under low dNTP conditions, and if dGTP were the sole A1 site activator, the pool levels would not readily reach the nanomolar regime observed in resting macrophages or CD4+ T cells. Such low dNTP pool levels have been implicated in the HIV-1 restriction activity of SAMHD1 in these cell types because reverse transcriptase is much less efficient when dNTP substrates are scarce (24, 25).

An enigmatic feature of SAMHD1 activity in resting immune cells is the presence of elevated dUTP levels in an otherwise depleted dNTP pool background (26, 27). Given our findings that dUTP is an excellent substrate, there must be as of yet undiscovered regulatory mechanisms that allow dUTP levels to persist or accumulate after the other dNTPs have been depleted. Since dUTP incorporation into HIV-1 cDNA has been implicated as a related antiviral defense mechanism (28, 29), a further understanding of the mechanisms that regulate nucleotide pools during the transition from active to resting states of cells will likely be relevant to viral infectivity and persistence in these cell types.
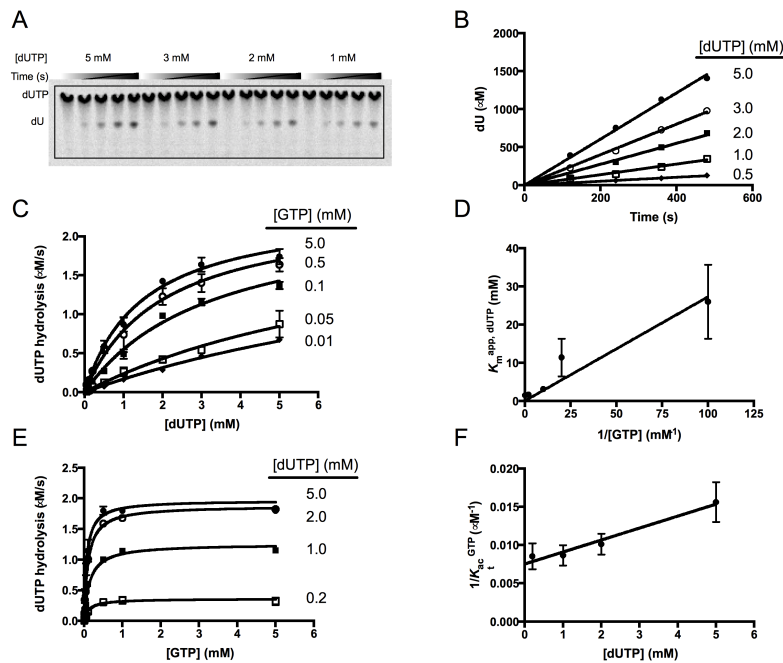
**Figure 2.1 Activation of dUTP hydrolysis by GTP.** Activation of dUTP hydrolysis by GTP. SAMHD1 (0.5 µM) was incubated with increasing concentrations of ^{3}H-dUTP in the presence of five GTP activator concentrations (0.01–5 mM). (*A*) Time courses for dUTP hydrolysis using the C18-RP TLC plate assay. (*B*) Linear initial rates for formation of the ^{3}H-dU product. (*C*) Initial velocities were found to follow a hyperbolic dependence on dUTP concentration with no evidence of sigmoidicity. Curves are from global non-linear least squares best fit to all of the data using eq 1. (*D*) Secondary plot of $K_m^{app, dUTP}$ against 1/[GTP]. (*E*) GTP concentration dependence of the dUTP hydrolysis rates at increasing [dUTP]. (*F*) Secondary plot of $1/K_{act}^{GTP}$ against [dUTP].
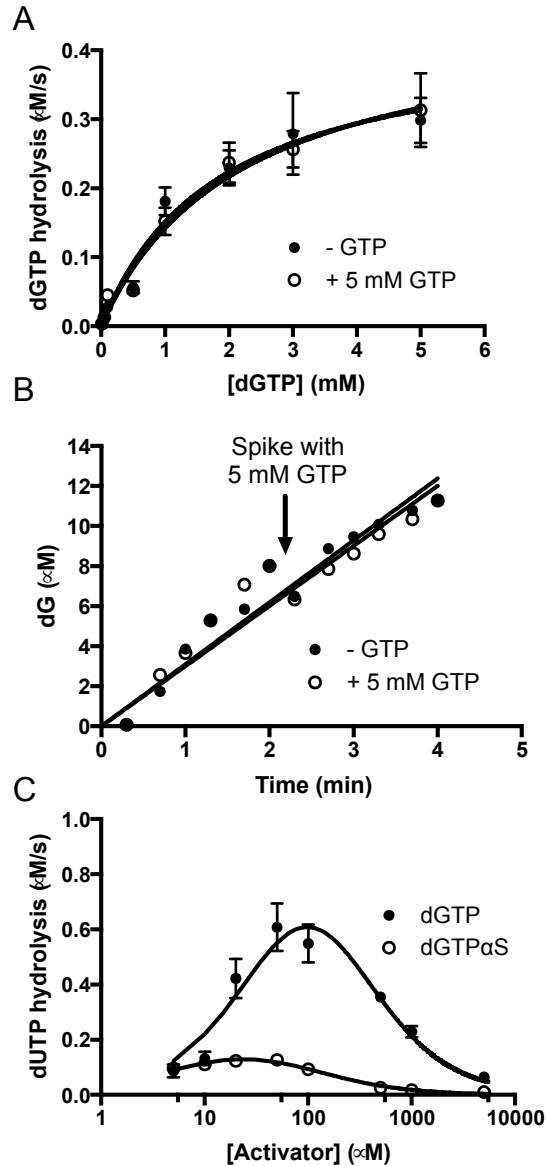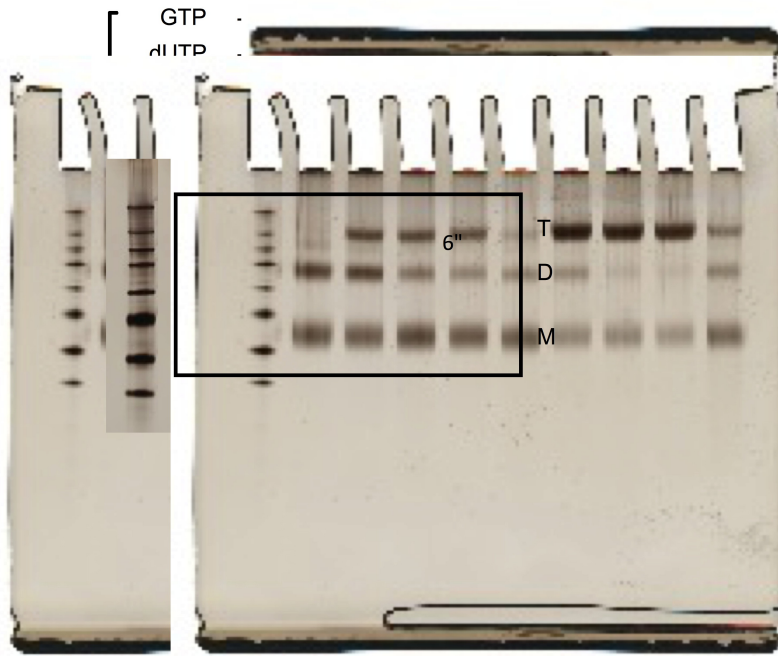
**Figure 2.2 Self-activation, transactivation, and inhibition by dGTP**. (*A*) dGTP self-activation of dGTP hydrolysis and trans-activation of dGTP hydrolysis by 5 mM GTP. The data sets were normalized to total active sites and are indistinguishable. (*B*) Addition of 5 mM GTP during steady-state hydrolysis of 25 μM dGTP by 0.5 μM SAMHD1 provides no additional activation. (*C*) Activation and inhibition of dUTP hydrolysis by dGTP and dGTPαS. Curves were generated using the simulation program Dynafit to obtain the activation and inhibition constants for dGTP (eqs 5-8, see text).
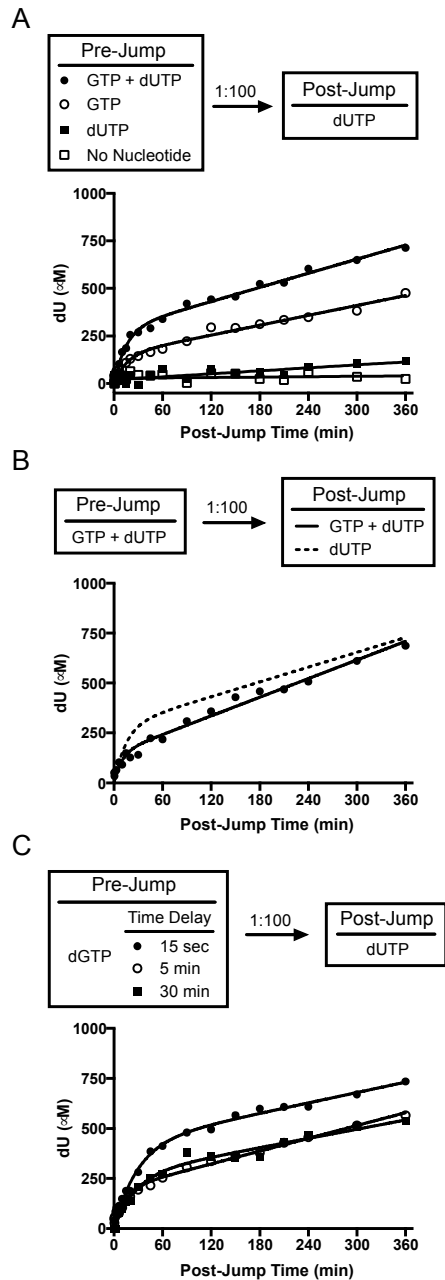
**Figure 2.3 Dilution-jump kinetic experiments for dUTP hydrolysis**. High concentrations of SAMHD1 and various nucleotide combinations were incubated as indicated in pre- and post-jump reactions. (*A*) SAMHD1 was incubated with GTP (0.5 mM) and dUTP (5 mM), GTP alone (0.5 mM), dUTP alone (5 mM), or no nucleotide prior to dilution. (*B*) Inclusion of 0.5 mM GTP in the post-jump reaction did not enhance the post-jump kinetics. For comparison, the dashed line shows the curve from the dUTP-only post-jump condition in panel (*A*). (*C*) Time dependence of pre-jump reactions containing only dGTP (5 mM).

A

GTP
dUTP

| + | + | + | + | + | + | - | - | - | - | - | - | - |
| - | + | + | + | + | + | - | - | - | - | - | - | - |
| - | - | - | - | - | - | + | + | + | - | - | - | - |
| - | - | - | - | - | - | - | - | - | + | + | + | + |
| + | - | - | + | + | + | + | + | - | - | + | + |

6"   T
     D
     M
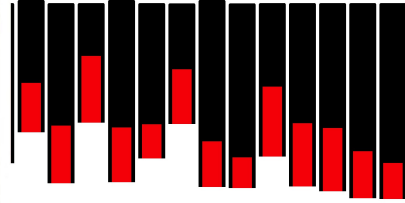
60  0  60  0  60 360  0  60 360  0  60  0  60

**Post-Jump Time (min)**

**for elucidating the oligomeric**
el separation of the cross-linked
cated pre-jump conditions prior
ate chemical cross-linking with
$P$ = 0.5 mM; dUTP, dGTP and
nucleotide concentrations were
ation was by silver staining and
d quantified using Quantity-One.
ns of monomer (M), dimer (D)
of dimer and tetramer stabilities
ns were performed as in (*A*) and
concentrations of pre- and post-
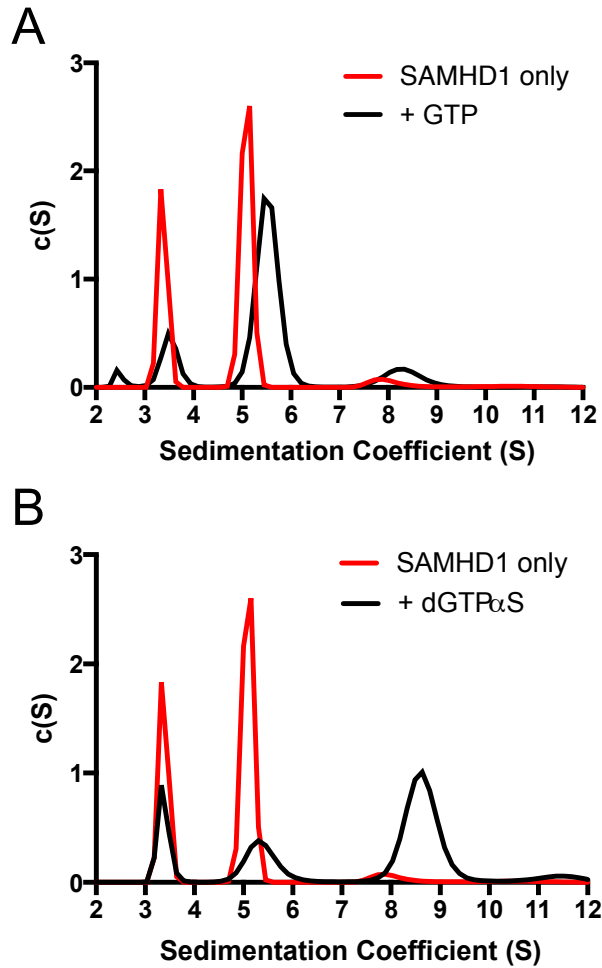
**Figure 2.5 Sedimentation velocity experiments to evaluate the effects of GTP and dGTPαS on the oligomeric state of SAMHD1**. (*A*) SAMHD1 alone (8 μM) and in the presence of 1 mM GTP or (*B*) SAMHD1 alone (8 μM) and in the presence of 1 mM dGTPαS were dialyzed and analyzed by sedimentation velocity centrifugation. Raw scan data were fitted to a c(*S*) distribution as described in the Methods.

**A**

| Nucleotide | M $\xrightarrow{K_D}$ D | D $\xrightarrow{K_T}$ T |
|---|---|---|
| - | 2.3 | ≥30 |
| GTP (1 mM) | 0.38 | 16 |
| dGTPαS (1 mM) | 0.41 | 0.89 |

**B**
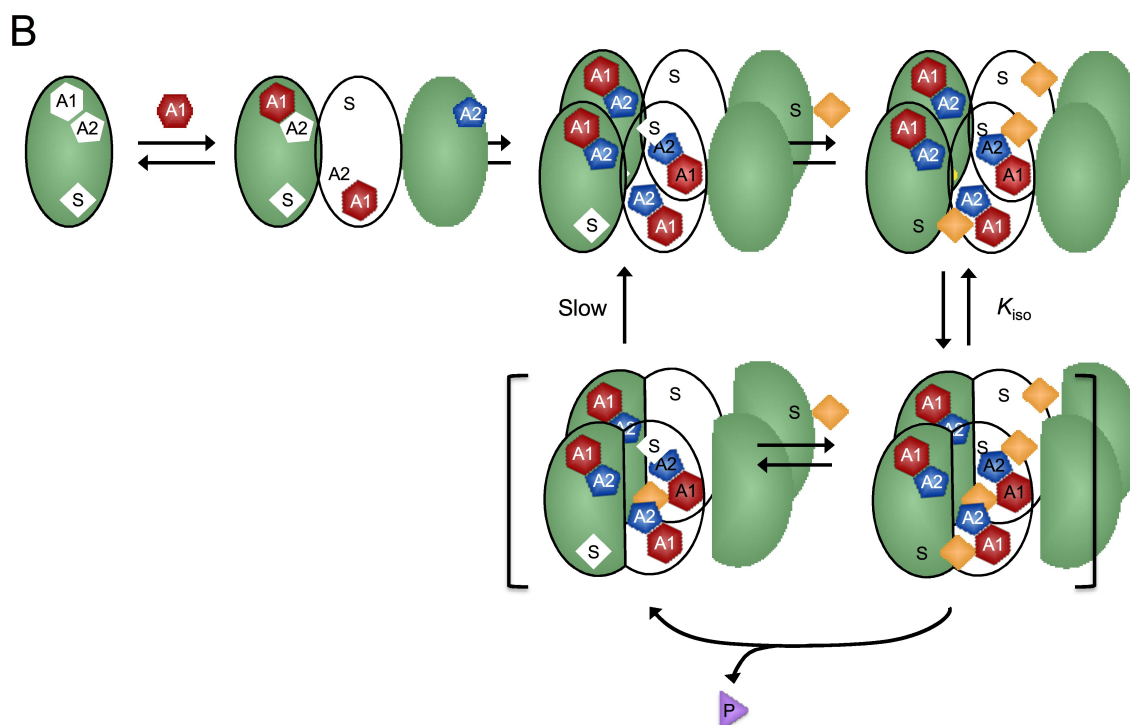
**Figure 2.6 Nucleotide-dependent oligomeric equilibria of SAMHD1 and mechanism of long-lived activation.** All values given are in concentration units of micromolar. (*A*) Summary of the effects of GTP and dGTPαS on the oligomeric equilibria. The values for $K_D$ and $K_T$ are for the indicated concentration of nucleotide. (*B*) Model for ordered-essential activation and oligomerization by GTP activators and dNTP co-activator/substrates involving the formation of a long-lived activated tetramer. The $E(A1A2)_4S_4$ complex is depicted as isomerizing ($K_{iso}$) to a long-lived tetramer (brackets) that performs steady-state turnover. This form only slowly reconverts to the loosely associated tetramer form that is in equilibrium with monomer, dimer and free nucleotides. The model also indicates that the activator sites of the long-lived tetramer do not communicate with free (d)GTPs during steady-state turnover of substrate dNTPs. Although not depicted, the actively cycling tetramer does not require the presence of bound activator nucleotides (see text).

58
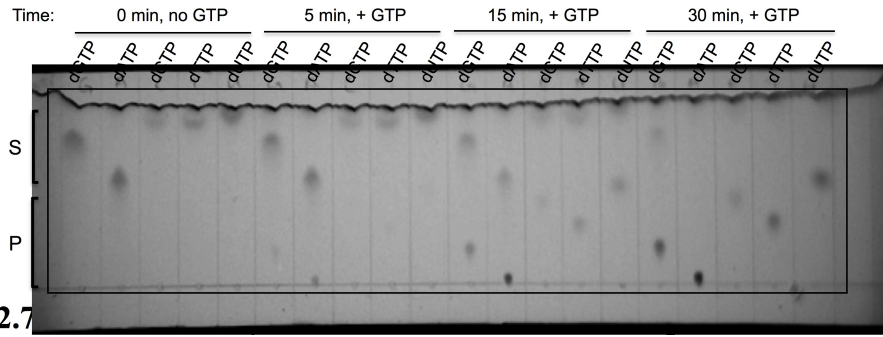
**Figure 2.7** [...] **containing protein 1 (SAMHD1).** Two microliter samples were removed at indicated times and quenched by spotting onto a FITC (UV$_{254}$) impregnated C18-reversed phase thin layer chromatography (TLC) plate. TLC plates were developed in 50 mM KH$_2$PO$_4$ (pH 4.0) to separate substrate from products. Plates were visualized using a Gel Doc (Bio-Rad) UV transilluminator (254 nm). The mobilities of the substrate dNTPs and product nucleosides are marked.
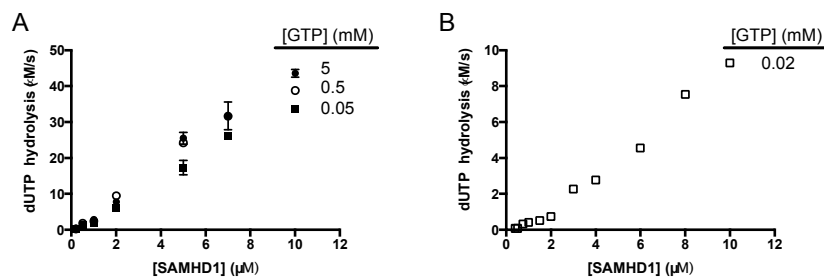
**Figure 2.8 Steady-state measurements of the rate of dUTP (1 mM) hydrolysis at various SAMHD1 concentrations.** (*A*) dUTP hydrolysis rate as a function of SAMHD1 monomer concentration at various concentrations of GTP activator. There is no evidence for significant nonlinear behavior in the range 50 ∝M to 5 mM [GTP]. (*B*) At very low concentrations of GTP (20 ∝M) and SAMHD1 the dUTP hydrolysis shows parabolic upward curvature indicative of changes in the active oligomeric state at this low concentration of activator. Such changes are expected based on the enzyme concentration dependence of the velocity sedimentation results (**Fig. S8**).
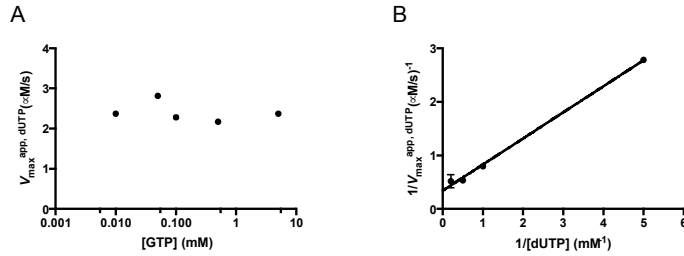
**Figure 2.9 Secondary replots to confirm the ordered essential activation mechanism.**
(A) The secondary plot of $V_{max}^{app, dUTP}$ against [GTP] activator concentration clearly demonstrates that dUTP substrate binding can always pull the equilibrium towards the active EAS complex at any concentration of activator. (*B*) Secondary plot $1/V_{max}^{app, dUTP}$ against 1/[dUTP] shows a linear response consistent with an ordered-essential activation mechanism.
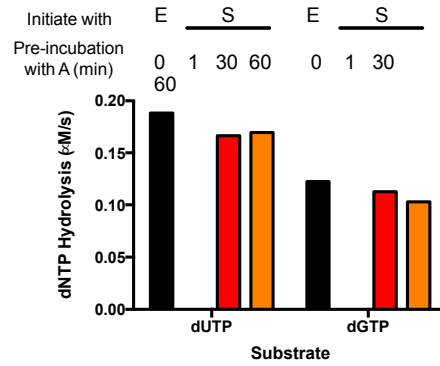
**Figure 2.10 The sequence of reagent addition shows no effect on the kinetic properties of SAMHD1**. Order-of-addition reactions contained GTP (A, 0.5 mM) trans-activator and either [5-^{3}H] dUTP (1 mM) or [8-^{3}H] dGTP (1 mM) substrate. Reactions were initiated by addition of SAMHD1 (0.5 $\mu$M) to samples containing GTP and [^{3}H] dNTP substrate (0 min). Alternatively, SAMHD1 was pre-incubated with GTP for the indicated time (1, 30 or 60 min) and then initiated by addition of substrate. Substrate hydrolysis was evaluated using the C18-RP TLC kinetic assay.
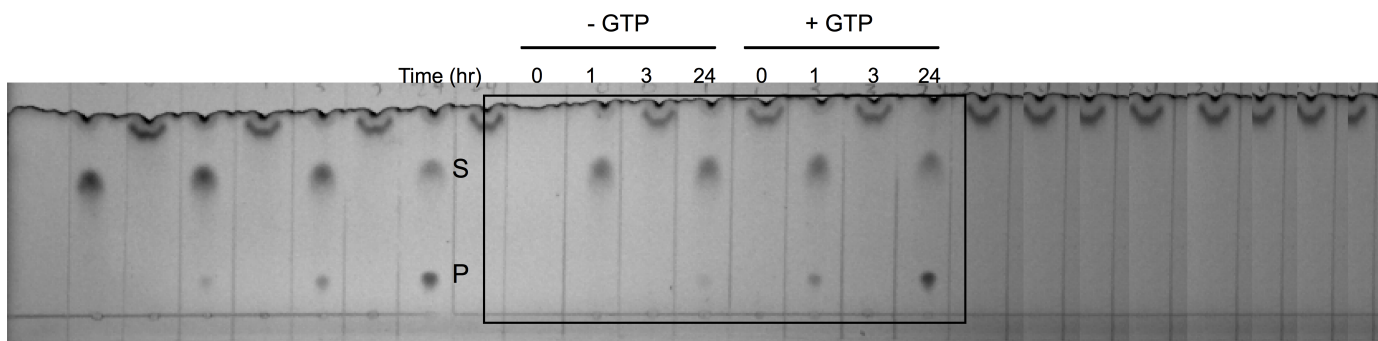
**Figure 2.11 Reactivity of 2'-deoxyguanosine-5'-[α-thio] triphosphate lithium salt.** (dGTPαS; 1 mM) with SAMHD1 (0.5 µM) in the presence or absence of GTP (50 µM). Reaction progress was monitored at 0, 1, 3 and 24 hr by spotting two microliter samples on C18-RP TLC (UV$_{254}$) plates. Results show that dGTPαS is a very slow substrate both in the absence and presence of GTP indicating self-activation.
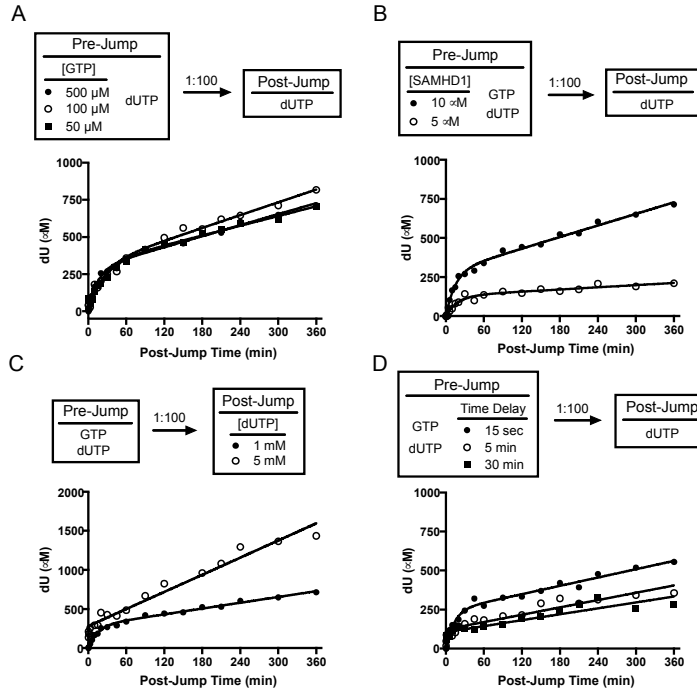
**Figure 2.12 Kinetic characteristics of various dilution-jump reactions.** (*A*) The burst amplitude was found to be independent of GTP concentration in a pre-jump solution that contained 50 to 500 ∝M GTP and 1 mM dUTP, indicating that the active enzyme form present in the post-jump burst-decay phase is completely assembled in the pre-jump time period over a wide range of GTP concentrations ($\leq$10 s). (*B*) The decay rate ($k_{inact}$) in post-jump reaction was independent of the SAMHD1 concentration in the pre-jump, but the burst amplitude and the rate of the second kinetic phase were directly proportional to the SAMHD1 concentration in the pre-jump. These findings established that the burst-decay phase is a uni-molecular decay process and that the second phase follows simple steady-state kinetic behavior. (*C*) The steady-state rate of the post-jump linear phase increased with the dUTP concentration present in the post-jump solution. (*D*) Delay-before-dilution experiments using GTP and dUTP in the pre-jump solutions established that as the substrate is depleted during the delay time before dilution, the burst amplitude in the post-jump reaction is reduced. Eventually, as the delay is increased to a length where all of the pre-jump substrate is consumed, the burst amplitude reaches the ~50% level observed when only GTP activator is present in the pre-jump solution. The similar results achieved with GTP alone in the pre-jump and using long pre-jump delay times with activator and substrate exclude the possibility that the long-lived active forms of SAMHD1 in the post-jump involve the PPP$_i$ and dN products generated in the pre-jump.
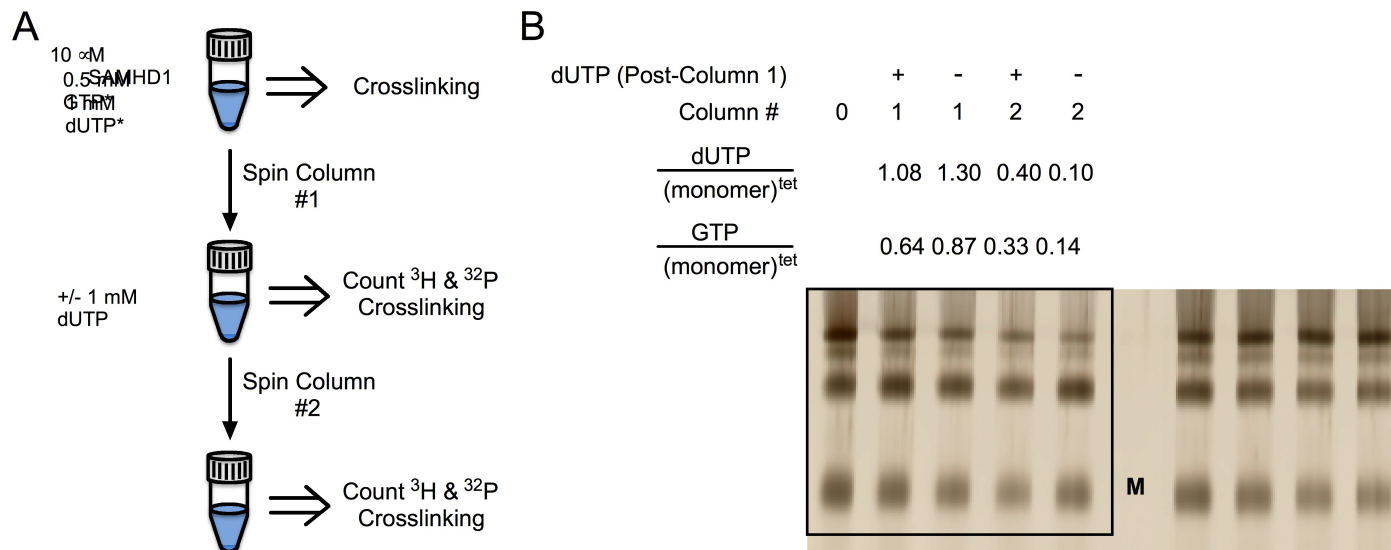
**Figure 2.13 Spin filtration assessment of activator nucleotide binding stoichiometry and release kinetics**. (A) Schematic of the experimental design showing the spin column steps to remove unbound nucleotides. GTP* and dUTP* represent the 32P- and 3 H-labeled nucleotides, respectively. In some experiments, the collection tube after the first spin column contained 1 mM unlabeled dUTP to mimic the postjump conditions of the dilution-jump protocol. (B) Denaturing PAGE was used to separate and quantify the cross-linked oligomers of SAMHD1 at the indicated step. The moles of GTP and dUTP present in the column excluded volume after each spin step were calculated from the known nucleotide-specific activity by scintillation counting. Because activator nucleotides are only expected to be tightly bound to the tetramer form, the stoichiometry is expressed as the ratio of nucleotide bound to each monomer of tetrameric SAMHD1 [(monomer)tet]. The amount of monomer was determined by quantifying the images of the gels, where the amount of monomer in the tetrameric form = (moles monomer loaded) × I tet/(I tet + I dim + I mon), and I tet, I dim, and I mon are the integrated intensities of the tetramer (T), dimer (D), and monomer (M) bands. The greater amounts of dUTP and GTP that are retained after the second spin when dUTP was present in the collection tube after the first spin show that cold dUTP is not accessible to the site where dUTP* is bound. This result is consistent with dUTP being bound to the sequestered A2 site. Because the second column was run 2 min after the first column, the half-life for nucleotide dissociation from the A1 and A2 sites is estimated to be around 2 min.
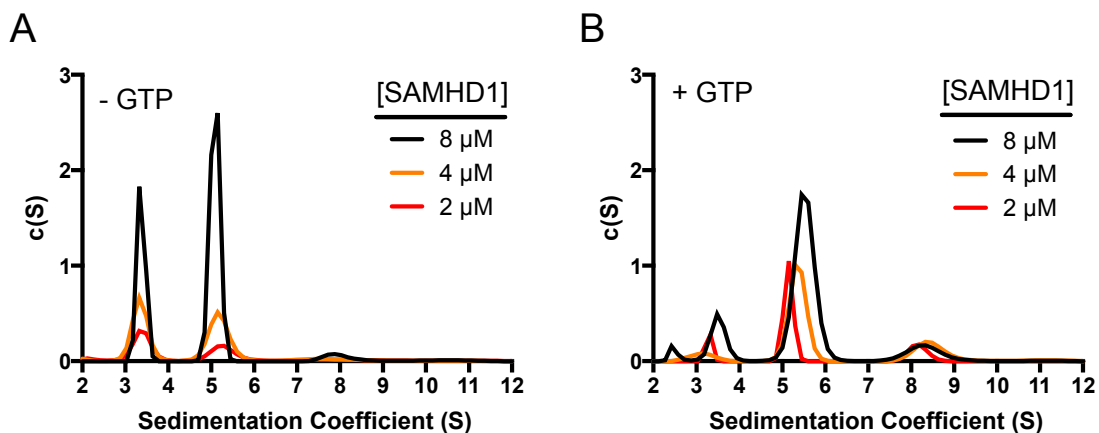
**Figure 2.14 Concentration dependence of SAMHD1 sedimentation velocity data**. The raw scan data were fitted to a c(*S*) distribution as described in the methods. (*A*) The indicated concentration of SAMHD1 was dialyzed and analyzed by sedimentation velocity. (*B*) The indicated concentration of SAMHD1 was dialyzed in the presence of GTP (1 mM) and analyzed by sedimentation velocity.
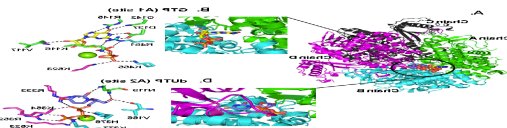
**Figure 2.15 Activator site interactions with GTP and** [...] P. (*A*) SAMHD1 tetramer colored by chain (chain A = green, chain B = cyan, chain [...] black, chain D = purple) with GTP and dUTP modeled in the A1 and A2 sites, res[...]vely (Mg$^{2+}$ shown as green sphere). (*B*) GTP binds to the Activator 1 site at th[...] monomer-monomer interface of chains A and B. Nucleobase-specific interactions are ma[...] th one monomer (chain A), while stabilization of the triphosphate moeity is accompl[...] primarily by residues from the second monomer (chain B). Introduction of the C3'-e[...] sugar pucker promotes the formation of a hydrogen bond between the 2'-OH of GT[...] the backbone carbonyl of V117, which is absent when C2'-endo dGTP is bound. (e[...] odeling of GTP at the A1 site (Mg$^{2+}$ omitted for clarity). Steric clash is introduce[...] he C3'-endo sugar pucker of GTP in the A2 site. The 2'-OH of the A2 site GTP is pos[...] d in close proximity to the 3'-OH of the GTP in A1 site relative to the C2'-endo sug[...] cker of dUTP as shown in panel B. (*D*) dUTP binds in the Activator 2 site at [...] portion of the dimer-dimer interface involving chains A, B and D. dUTP is stabilize[...] narily by extensive contact with chains B and D. The reduced number of base-speci[...] ntacts in A2 relative to A1 implies nucleobase promiscuity in the A2 site. Exten[...] contacts established at the dimer-dimer interface suggest the Activator 2 sit[...] fully formed only upon tetramerization of SAMHD1. Residues are color-code[...] chain based on (A). (*E*) Modeling of dUTP at the A2 site (Mg$^{2+}$ omitted for cla[...] Space filling models of the sugar moieties show acceptable fit with no overlap of va[...] Waals radii of dUTP (blue) in the Activator 2 site when the Activator 1 site is occupi[...] GTP (yellow).
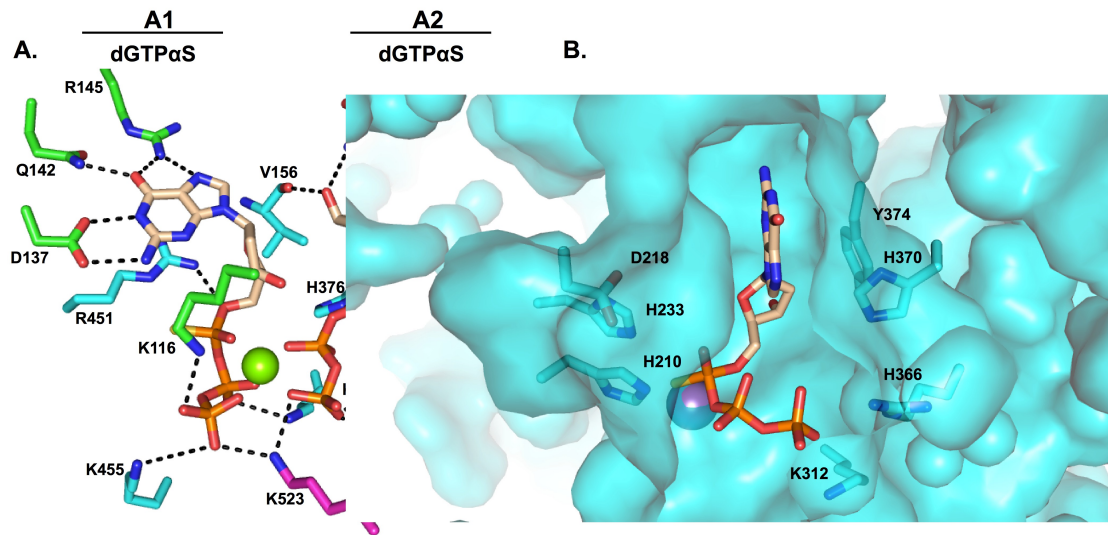
**Figure 2.16 Interactions of dGTPαS in activator and catalytic sites.** (*A*) Binding of dGTPαS in the A1 site is stabilized primarily by the monomer-monomer interface between chains A and B. dGTP αS in the Activator 2 site makes extensive contact at the dimer-dimer interface. Occupation of both A1 and A2 sites is facilitated by contacts between the ligands and the tetrameric form of SAMHD1. (*B*) Tetrameric SAMHD1 produces a compact, catalytic site capable of accommodating all dNTPs, including dGTPαS (pink). Essential contact and catalytic residues are highlighted and $Mn^{2+}$ is shown as a purple sphere. Residues are color-coded by chain based on Figure S13A. (PDB ID = 4BZC)

# References

1. Yao NY, Schroeder JW, Yurieva O, Simmons LA, O'Donnell ME (2013) Cost of rNTP/dNTP pool imbalance at the replication fork. *Proc Natl Acad Sci USA* 110(32):12942-12947.
2. Bebenek K, Roberts JD, Kunkel TA (1991) The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *J Biol Chem* 267(6):3589-3596.
3. Gavegnano C, Kennedy EM, Kim B, Schinazi RF (2012) The impact of macrophage nucleotide pools on HIV-1 reverse transcription, viral replication, and the development of novel antiviral agents. *Mol Biol Int* 2012(5270):1-8.
4. Lahouassa H, et al. (2012) SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* 13(3):223-228.
5. Franzolin E, et al. (2013) The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc Natl Acad Sci USA* 110(35):14272-14277.
6. Hakansson P, Hofer A, Thelander L (2006) regulation of mammalian ribonucleotide reduction and dNTP pools after DNA damage and in resting cells. *J Biol Chem* 281(12):7834-7841.
7. Kashlan OB, Scott CP, Lear JD, Cooperman BS (2002) A comprehensive model for the allosteric regulation of mammalian ribonucleotide reductase. Functional consequences of ATP- and dATP-induced oligomerization of the large subunit. *Biochemistry* 41(2):462-474.
8. Engström Y, et al. (1985) Cell cycle-dependent expression of mammalian ribonucleotide reductase. *J Biol Chem* 260(16):9114-9116.
9. Rampazzo C, et al. (2010) Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutat Res Int* 703:2–10.
10. Cribier A, Descours B, Valadão ALC, Laguette N, Benkirane M (2013) Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. *Cell Rep* 3(4):1036-1043.
11. Li N, Zhang W, Cao X (2000) Identification of human homologue of mouse IFN-γ induced protein from human dendritic cells. *Immunol Lett* 74(3):221-224.
12. Vorontsov II, et al. (2011) Characterization of the deoxynucleotide triphosphate triphosphohydrolase (dNTPase) activity of the EF1143 protein from *Enterococcus faecalis* and crystal structure of the activator-substrate complex. *J Biol Chem* 286(38):33158-33166.
13. Kornberg SR, Lehman IR, Bessman MJ, Simms ES, Kornberg A (1958) Enzymatic cleavage of deoxyguanosine triphosphate to deoxyguanosine and tripolyphosphate. *J Biol Chem* 233(1):159–162.
14. Kohn G, et al. (2012) High inorganic triphosphatase activities in bacteria and mammalian cells: Identification of the enzymes involved. *PLoS ONE* 7:e43879.
15. Goldstone DC, et al. (2011) HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* 480(379):379-382.
16. Kim ET, White TE, Brandariz-Nunez A, Diaz-Griffero F, Weitzman MD (2013) SAMHD1 restricts herpes simplex virus type 1 (HSV-1) in macrophages by limiting DNA replication. *J Virol* 87(23):12949–12956.

17. Powell RD, Holland PJ, Hollis T, Perrino FW (2011) Aicardi-Goutieres syndrome gene and HIV-1 restriction factor SAMHD1 is a dGTP-regulated deoxynucleotide triphosphohydrolase. *J Biol Chem* 286(51):43596-43600.
18. Baldauf H-M, et al. (2012) SAMHD1 restricts HIV-1 infection in resting CD4+ T cells. *Nat Med* 18(11):1682-1687.
19. Amie SM, Bambara RA, Kim B (2013) GTP is the primary activator of the anti-HIV restriction factor SAMHD1. *J Biol Chem* 288(35):25001-25006.
20. Ji X, et al. (2013) Mechanism of allosteric activation of SAMHD1 by dGTP. *Nat Struct Mol Biol* 20(11):1304-1309.
21. Yan J, et al. (2013) Tetramerization of SAMHD1 is required for biological activity and inhibition of HIV infection. *J Biol Chem* 288(15):10406–10417.
22. Zhu C, et al. (2013) Structural insight into dGTP-dependent activation of tetrameric SAMHD1 deoxynucleoside triphosphate triphosphohydrolase. *Nat Commun* 4:2722.
23. Segel IH (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems* (Wiley, New York), pp 227-272.
24. Diamond TL, et al. (2004) Macrophage tropism of HIV-1 depends on efficient cellular dNTP utilization by reverse transcriptase. *J Biol Chem* 279:51545–51553.
25. O'Brien WA, et al. (1994) Kinetics of human immunodeficiency virus type 1 reverse transcription in blood mononuclear phagocytes are slowed by limitations of nucleotide precursors. *J Virol* 68(2):1–7.
26. Kennedy EM, et al. (2011) Abundant non-canonical dUTP found in primary human macrophages drives its frequent incorporation by HIV-1 reverse transcriptase. *J Biol Chem* 286(28):25047–25055.
27. Aquaro S, Calio R, Balzarini J, Bellocchi MC, Garaci E (2002) Macrophages and HIV infection: therapeutical approaches toward this strategic virus reservoir. *Antiviral Res* 55:209-225.
28. Weil AF, et al. (2013) Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc Natl Acad Sci USA*:e448.
29. Priet S, Sire J, Querat G (2006) Uracils as a cellular weapon against viruses and mechanisms of viral escape. *Curr HIV Res* 4(1):31–42.
30. Kuzmic P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem* 237(2):260–273.
31. Schuck P, Rossmanith P (2000) Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* 54(5); 328-341.
32. Stafford WF, Sherwood PJ (2004) Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. *Biophys Chem* 108(1):231-243.
33. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1991) *Analytical Ultracentrifugation in Biochemistry and Polymer Science* eds Harding SE, Rowe AJ, Horton JC (Royal Society of Chemistry, Cambridge), pp 90-125.

# Chapter 3: Diverse fates of uracilated HIV-1 DNA during infection of myeloid lineage cells

**Abstract:** We report that a major subpopulation of monocyte-derived macrophages (MDMs) has a metabolic phenotype leading to high levels of dUTP incorporation into HIV-1 DNA during reverse transcription (U/A pairs). Uracilated viral DNA products are degraded by the nuclear uracil base excision repair (UBER) machinery with less than 1% of the uracilated DNA successfully integrating. Although uracilated proviruses showed relatively few mutations, the viral genomic RNA was highly mutated, suggesting error-prone transcription and/or repair. Despite the restrictive MDM environment, infectious virus was still produced. Viral DNA isolated from blood monocytes and alveolar macrophages (but not T cells) of drug-suppressed HIV-infected donors also contained detectable uracils. The presence of viral uracils in short-lived circulating monocytes suggests their recent infection through contact with virus producing cells in a tissue reservoir. These findings characterize a new viral defense mechanism involving host UBER that is relevant to the establishment and persistence of HIV-1 infection.

## Introduction

The uracil nucleobase plays a central role in adaptive and innate immunity against HIV-1 when it is found in DNA rather than RNA [1,2]. The most well characterized uracil-centric innate immune response involves host cell DNA cytidine deaminase enzymes (APOBECs), which selectively deaminate cytosine residues during (-) strand DNA synthesis thereby rendering the viral genome nonfunctional by hypermutation (C/G→ T/A). Work from our lab has suggested the presence of another uracil-mediated

HIV-1 restriction pathway involving the incorporation of dUTP into viral DNA by reverse transcriptase to produce U/A base pairs ("uracilation")[3]. The dUTP-dependent pathway is thought to be restricted to non-dividing immune cells such as macrophages, monocytes and resting CD4+ T cells because only non-dividing cells have the requisite low levels of canonical dNTPs and elevated ratios of dUTP/TTP [4,5]. Notably, U/A pairs resulting from dUTP incorporation are "invisible" to normal DNA sequencing methods and they retain the coding potential of normal T/A base pairs. Despite the similarity of U/A and T/A pairs, the presence of uracil in DNA has the potential to introduce diverse effects on viral infection including transcriptional silencing and engagement of the host uracil base excision repair (UBER) pathway [6-8]. The presence, persistence, and ultimate fate of U/A pairs in HIV-1 proviral DNA is unexplored and is important for understanding the potential of tissue macrophages to establish and maintain a long-term HIV reservoir.

**Results**

**Uracilation of HIV DNA results from the unique nucleotide metabolism in myeloid phagocytic cells.** We hypothesized that viral uracilation and restriction in resting immune cells would require enzyme activities that support a high dUTP/TTP ratio and uracil base excision and DNA fragmentation. Using sensitive and specific *in vitro* enzymatic assays (**Fig. S1A-D**)[3,9,10], we found that monocytes and monocyte-derived macrophages (MDMs) expressed high levels of SAMHD1 dNTP triphosphohydrolase to reduce the canonical dNTP pools[9,11], undetectable dUTPase activity that allowed dUTP accumulation, and modest expression of the UBER enzymes uracil DNA glycosylase

(hUNG) and abasic site endonuclease (APE1) (**Fig. S1E-H**). Although resting CD4⁺T

cells also possessed high SAMHD1, hUNG and APE activities, their dUTPase activity

was at least 7-fold greater than MDMs. LC-MS analyses of the dUTP and canonical

dNTP levels in resting and activated $CD4^+$ T cells and MDMs revealed that the

dUTP/TTP ratio was ~20 for MDMs, 1.1 for resting $CD4^+$ T cells, and <0.05 for

activated $CD4^+$ T cells (**Fig. S1I, J**)[12,13].

**Total HIV-1 DNA isolated from in vitro infected MDMs is heavily uracilated.** To

detect and map uracil across the HIV genome we developed uracil excision-digital

droplet PCR (Ex-ddPCR) (**Fig. 1A**)**.** Briefly, Ex-ddPCR involves isolation of total DNA

from HIV infected MDMs, after which half of the sample is treated with UNG to destroy

template strands that contain uracil (**Fig. 1A**). Thus, any PCR amplicon that contains one

or more uracils on each DNA strand is not amplified in the UNG treated DNA sample.

After performing ddPCR (no UNG treatment) and Ex-ddPCR (UNG pre-treatment), the

fraction of the amplicons containing at least one uracil on each strand was calculated

from the counts of positive droplets for the ddPCR and Ex-ddPCR samples (**Fig. 1A**). A

complete description of Ex-ddPCR is found in the **Supplementary Information** and **Fig.**

**S2A-E**.

Ex-ddPCR analysis of HIV DNA isolated from single-round infections of activated

and resting CD4+ T cells and MDMs using a VSVG pseudo-typed HIV strain (HIV $^{NL4.3}$)

established that uracilation of viral DNA occurs in MDMs, but not T cells (**Fig. 1B**). The

scatter plots and histograms in Figure 1B show that the copy number for viral *gag* DNA

isolated from MDMs at 3 days post-infection was 3.5-fold lower in the Ex-ddPCR

experiment, showing that ~70% of *gag* amplicons contained uracil. Importantly, the

genomic reference standard RNase P (*RPP30*) contained no measurable uracil, indicating that uracil incorporation is specific to HIV DNA. In contrast, the *gag* copy number for viral DNA collected from activated and resting T cells was the same for both the ddPCR and Ex-ddPCR reactions indicating that uracil was absent in viral DNA isolated from infected T cells (**Fig. 1B**).

To augment Ex-ddPCR, we also applied the next generation sequencing technology uracil Ex-Seq to globally map the frequency of U/A pairs across the entire HIV genome[14]. Ex-Seq is similar to standard IIlumina sequencing (Seq), except that UNG-mediated uracil excision is used to destroy uracil-containing templates prior to PCR amplification. To specifically enrich HIV-1 sequences, we used 5'-biotin conjugated DNA probes that tiled both strands of the entire viral genome, yielding a $10^3$-fold increase in HIV-derived fragments. Sequencing of viral DNA isolated from MDMs seven days after infection with HIV[NL4.3] showed uniform coverage across the genome except for notably increased reads at the 5′ and 3′-LTR regions, which was equally evident for both the Seq and Ex-Seq samples (**Fig. 1C**). We speculate that the elevated signal in the LTRs arises from non-uniform hybridization of the lock-down probes, which is normalized when converting the reads to Frac U. The ratio of the normalized sequencing reads (Ex-Seq/Seq) thus quantifies the fraction of sequence reads that contained at least one uracil on each strand (**Fig. 1D**). This ratio indicates that on average about 60% of the 100 bp reads contained uracil and uracilation was fairly uniform across the genome.

**Uracils do not arise from APOBEC-catalyzed cytosine deamination.** MDMs have previously been shown to have little if any APOBEC DNA cytidine deaminase activity [15],

which, if present, would introduce uracils via a parallel cytidine deamination pathway. We confirmed that APOBEC activity was not important in control experiments described in the **Supplemental Information** (**Fig. S3A-F)**. We also established that the level of viral DNA uracilation could be greatly attenuated by increasing the intracellular TTP levels by adding thymidine to the culture media (**Fig. S3G-I**). These results are consistent with our proposal that the uracils arise from dUTP utilization by reverse transcriptase.

**Integrated proviruses contain abundant uracils.** The above analyses report on both integrated and unintegrated viral DNA. However, the Ex-Seq experiment also provides specific information on the uracilation status of integrated proviruses and their genomic sites of integration (**Fig. 1E)**. Sequence reads from integrated proviruses are unambiguously assigned by the presence of viral LTR and human genomic sequences ("discordant pairs"). Over 4000 discordant pairs were identified per million HIV reads for the Seq sample, which was reduced by about 60% for the Ex-Seq sample (**Fig. 1E)**. This level of uracilation is similar to that of the total HIV DNA and reinforces our previous assertion that uracils can persist in proviruses when nuclear hUNG2 expression is low[3].

**Identification of restrictive and permissive MDM subpopulations.** During HIV infection of MDMs ~10% of the infected MDMs were positive for GFP fluorescence after a single-round infection with HIV[NL4.3] (**Fig. 2A)**, which remained constant from about 7 to 30 days after infection and did not vary when the MOI was varied in the range 0.1-10 (**Fig. S4)**. This result led us to suspect that the MDM population consisted of a phenotypic mixture with different susceptibilities to HIV infection. (We note that higher

MOIs are typically used in our experiments to obtain reasonable viral copy numbers in the GFP$^-$ population). Accordingly, we used GFP fluorescence to sort HIV$^{NL4.3}$ infected MDMs into 99% pure GFP positive and negative populations (**Fig. S5**). Digital droplet PCR measurements showed that the copy number of early viral DNA products in the GFP$^-$ population (detected using a *gag*-specific PCR primer) was about 7-fold lower than the GFP$^+$ population at one-day post infection (~2 versus 14 copies/cell, **Fig. 2B**). This result suggested that viral infection was hindered at the entry and/or early reverse transcription steps in the GFP$^-$ population. Using Alu-*gag* nested qPCR to measure proviral copies[16], we found that the GFP population contained only 1% of the copies seen in the GFP$^+$ MDMs (**Fig. 2C**). The large decrease in copy number between the early DNA intermediates and the provirus stage indicates that a potent pre-integration restriction mechanism is present exclusively in the GFP$^-$ population.

**Viral uracilation is exclusive to GFP$^-$ MDMs.** We used infections with a replication-competent virus HIV$^{SF162}$ to further elucidate the characteristics of the two MDM populations. First, the infected MDMs were sorted into GFP positive and negative populations at one-day post infection. The cells were then cultured for 30 days and at one, seven, 14, and 30 days post-sorting we measured proviral copy numbers, the fraction of uracil-containing proviruses, and the levels of viral growth by ELISA for HIV-1 p24 in the culture supernatants for each cell population (**Fig. 2D, E**). Both the GFP positive and negative populations showed high levels of viral *gag* DNA products at day one (red circle and blue square, **Fig. 2D**). The GFP$^+$ population was characterized by efficient integration of viral DNA products and a high proviral copy number that remained stable for 14 days, before increasing by ~4-fold, which we attribute to secondary infection by

77

released viruses (white bars, **Fig. 2D**). In contrast, the GFP⁻ population had a low efficiency of integration exemplified by a 50-fold reduction in copy number between the early *gag* DNA products and proviruses (black bars, **Fig. 2D**). Unlike the GFP⁺ cells, the copy number in the GFP⁻ population decreased 20-fold between days seven and 14 before returning to the day seven level at 30 days. We attributed the increase to secondary infection by released virus particles because it was not observed in single-round infections. We established that HIV^NL4.3, HIV^SF162, and HIV^BAL viral strains and MDMs isolated from different donors behaved similarly (**Fig. S6A-E**), and that the results were independent of whether differentiated MDMs were obtained using adherence, M-CSF, or GM-CSF protocols (**Fig. S6F-H**).

The most striking difference between the two MDM populations was that the minor GFP⁺ population had no detectable proviral uracils, while ~90% of the proviral copies in the major GFP⁻ fraction contained uracil at day one (**Fig. 2D**). The fraction of proviral copies that were uracilated (Frac U) decreased from 0.77 to 0.23 between days 7 and 14 before rising to 0.55 at day 30 (presumably due to continuing infection). We explored the mechanistic basis for the differences in viral uracilation between the two MDM populations by measuring dUTPase, hUNG2, and SAMHD1 activities and the dUTP/TTP ratio in the cell extracts (**Table S1**). The only salient difference in these parameters was the 20-fold lower dUTP/TTP ratio in the GFP⁻ MDMs, which could partially or fully account for the absence of uracils in these cells.

Consistent with their larger proviral copy numbers, the GFP⁺ MDMs produced 100 to 1000 times more p24 than the GFP⁻ MDMs and the levels were fairly constant over a 30-day period (**Fig. 2E**). In contrast, the GFP⁻ MDMs showed a 10-fold decrease in p24

78

levels between day 7 and 14 (**Fig. 2E**), which coincides with the 20-fold reduction in proviral copy number in the same time period (**Fig. 2D**). The expression of detectable levels of p24, but not GFP, in the GFP $^-$ MDMs suggests that truncated viral RNAs are being expressed that extend through the p24 coding sequence, but not GFP.

**Viral uracilation is modulated by hUNG2 activity.** We previously reported that in the HT29 model cell system nuclear hUNG2 activity was the primary determinant of whether uracilated viral DNA survived or persisted in cells [3]. To confirm this result in MDMs we took advantage of viral protein R (vpr), which interacts with hUNG2 and induces its degradation through the assembly with the DDB1-CUL4 ubiquitin ligase complex [17]. We surmised that infection of MDMs with a viral construct without vpr (HIV$^{NL4.3}$Δvpr) would lead to increased hUNG2 activity, fewer proviral copies, and lower uracil levels in proviral DNA.

We determined that the nuclear (hUNG2) and mitochondrial (hUNG1) isoforms were detectable by western blotting (**Fig. S8**) and that HIV $^{NL4.3}$Δvpr virus did not express vpr (**Fig. 3A**). Infections of MDMs with HIV$^{SF162}$ and HIV$^{NL4.3}$Δvpr showed that between one and seven days post-infection the hUNG2 nuclear isoform selectively disappeared with HIV (vpr$^+$), but not HIV$^{NL4.3}$Δvpr (**Fig. 3A**)[17]. As expected, infection with HIV$^{NL4.3}$Δvpr reduced the overall proviral copies present in the mixed population of GFP$^+$ and GFP$^-$ MDMs (**Fig. 3B**), which can be attributed to any of the known pro-infective functions of vpr[17]. The most instructive result was that in the absence of vpr fewer proviruses contained uracil (Frac U decreased by 2-fold) (**Fig. 3C**). This result

specifically reports on vpr-induced depletion of hUNG2 activity in the GFP MDMs because uracils are only present in GFP ⁻ MDMs and hUNG is the only glycosylase that acts on U/A base pairs in DNA[18].

**Uracilated proviruses produce fewer viral genomic RNAs.** We investigated whether the presence of uracil in proviral DNA affected the copy number of extracellular viral genomic RNAs (EVRs) present in supernatants of infected MDMs cultured in standard growth media or media supplemented with IFNγ or IL-4 (**Fig. 4A**). The EVR copies present in the supernatants of GFP$^+$ and GFP$^-$ MDMs were measured using reverse transcriptase quantitative PCR (RT-qPCR) at three days post-cytokine stimulation. These cytokines were of interest because they stimulate macrophage differentiation into classical pro-inflammatory (M1) and anti-inflammatory (M2) polarization states, respectively[19]. We found that EVR copies were about 100-fold lower for the GFP ⁻ population than the GFP$^+$ MDMs, regardless of cytokine stimulation and that IFNγ stimulation reduced the EVR copy number in both MDM populations (**Fig. 4B**).

**Extracellular viral RNAs (EVRs) have more mutations than uracilated proviral DNA.** Limiting dilution clonal sequencing was performed on EVRs produced from sorted GFP$^-$ and GFP$^+$ MDMs at three days post-stimulation[20,21], revealing surprising differences in viral single-point mutation frequencies for these populations (**Fig. 4C**). Regardless of cytokine stimulation, the GFP $^+$ population showed no detectable mutations in the roughly 12,000 total bases of the V3/V4 variable region of *env* that were sequenced. In contrast, the GFP population showed a 1.2% basal mutation frequency, which responded differently to IFNγ and IL4 stimulation:  IFNγ stimulation increased the

frequency by 2-fold and IL4 decreased it significantly (**Fig. 4C**). These results suggest

that the cytokine environment of an *in vivo* infected macrophage will influence the fate of

uracilated proviruses. In contrast to the mutated *env* sequences, the LTR region of EVRs

produced from the GFP⁻ cells was devoid of detectable mutations after sequencing 7,392

total bases (**Table S2**), which may reflect selection for expression of transcription

competent LTR sequences. The absence of mutations in the *env* and LTR regions of

EVRs isolated from the GFP⁺ MDM cultures, as well as the LTR region of EVRs

obtained from the GFP⁻ culture supernatants, virtually eliminates the possibility that the

observed mutations in *env* arise from PCR or sequencing errors.

The frequencies of single-base transition, transversion, and insertion/deletion

mutations within the *env* region of EVRs produced from IFN, IL4 and unstimulated

GFP⁻ cells are shown in **Fig. 4C**. Analysis of the observed G→A (+) strand mutations

using Hypermut 2.0 [22] indicated that the sequence dependence and frequencies were not

attributable to APOBEC deaminase activity on the viral (-) strand cDNA[23]. The

remaining RNA (+) strand mutations were primarily single base changes mostly

comprised of A→G, A→C, and T→C transitions and transversions (**Fig. 4D,E**), but also

a few short 2-3 nucleotide insertions or deletions.

Targeted sequencing of the *env* amplicon in uracilated viral DNA isolated from

GFP⁻ MDMs at day fourteen post-infection in the absence of cytokine stimulation

revealed a <0.007% mutation frequency in the V3/V4 variable region, which may be

contrasted with the 1.2% frequency in EVRs produced from the same cell culture (most

viral DNA is integrated at day fourteen). This result suggests that the EVR mutations

arise from transcriptional errors by RNA polymerase II (RNAP II). Such mutations may

derive from encounter of the polymerase with either uracils or repair intermediates resulting from uracil excision (see Discussion). Illumina next-generation sequencing data covering the entire viral DNA genome (obtained from total HIV $^{NL4.3}$ viral DNA isolated from a mixed population of MDMs at day seven post-infection in the absence of cytokines) confirmed the low mutation frequency determined from targeted sequencing of the *env* V3/V4 variable region (see Data Access information).

To explore whether the mutation signature of uracilated viral DNA products and viral genomic RNA sequences derived from infected MDMs were distinct from T cells, we infected activated CD4+ T cells with HIV$^{BAL}$ virions produced from donor cells expressing endogenous levels of APOBEC3G (A3G) (**Supplemental Methods**). The infection was kept to a single round by inclusion of the entry inhibitor enfuvirtide [24], and the V3/V4 *env* region of 14 proviruses were sequenced. The proviral sequences derived from infected T cells had a much higher mutation frequency (~1%) than uracilated viral DNA in MDMs (<0.007%), but the frequency was similar to the EVRs expressed by infected GFP$^-$ MDMs cultured in the absence of cytokines (1.2%, **Fig. 4C**).

The mutation spectrum of proviral DNA isolated from T cells differed from the EVRs produced from MDMs (**Fig. 4E**). The distinguishing proviral mutations found in T cells are elevated C→T transitions and T→G transversions, while the viral RNA sequences derived from MDMs are distinguished by elevated →C transversions and T→C transitions (**Fig. 4E**).  We note that previous studies have established that host RNAP II contributes little to HIV sequence evolution in T cell infections and that errors during reverse transcription are the predominant diversification mechanism [25]. This does not appear to hold for *in vitro* infection of MDMs.

**Uracilated proviruses produce replication competent HIV.** To measure the levels of functional progeny viruses that emerged from infected MDMs in the absence and presence of cytokine stimulation we followed the approach outlined in **Figure 5A**. MDMs infected with HIV [SF162] were first sorted based on GFP fluorescence at one-day post-infection. The sorted GFP positive and negative cells were divided into three cultures that were stimulated with IFN γ, IL4, or no treatment. After three days the MDM producer cultures were analyzed with respect to (i) *gag* DNA copy number as a surrogate for total viral DNA present (**Fig. 5B**), (ii) integrated proviral DNA copy number (**Fig. 5C**), (iii) the fraction of proviruses containing uracil (**Fig. 5D**), and (iv) p24 levels in culture supernatants as a surrogate for progeny virus output (**Fig. 5E**). The virus-containing MDM culture supernatants were removed at three days post-stimulation (dps), normalized to p24 and added to cultures of CEMx174 target cells. After three days, the proviral copy number per $10^6$ target cells was measured (**Fig. 5F**).

Independent of stimulation, the GFP⁻ MDM producer cells showed lower copy numbers of *gag* and proviral DNA (~20 and $10^4$-fold) and p24 levels (100-fold), as compared to GFP⁺ MDMs[26]. IFNγ or IL4 stimulation had little effect on *gag* and proviral DNA copy numbers in both MDM producer populations (**Fig. 5B,C**), but IL4 induced a modest decrease in the fraction of proviruses that contained uracil at three days post stimulation (**Fig. 5D**). For both MDM types, the addition of either cytokine resulted in a reduction in p24 production, with IFNγ showing the largest effect relative to no stimulation (~10-fold decrease) (**Fig. 5E,F**). Despite the much lower levels of virus

83

produced from GFP⁻ MDM producer cells, the harvested viruses derived from uracilated proviruses were capable of infecting CEMx174 target cells (**Fig. 5G**).

**Uracils are present in HIV-1 DNA from peripheral blood monocytes of virally suppressed donors.** To test whether HIV-1 DNA isolated from infected individuals contained detectable levels of uracil we purified resting CD4 T cells and monocytes from bulk peripheral blood mononuclear cells (PBMCs) of six participants suppressed on antiretroviral therapy (ART). Highly purified populations of resting CD4+ T (>95%) cells and monocytes (>98%) were obtained and quantified by flow cytometry **(Fig. S8A-C)**. ddPCR and Ex-ddPCR were used to quantify HIV DNA and measure the fraction of the viral *pol* amplicons that contained uracil **(Fig. 6A,B)** [27]. HIV DNA levels in resting CD4+ T cells were generally higher and more variable than those measured in monocytes (geometric mean values of 526 copies/10 T cells and 47 copies/10 monocytes) **(Fig. 6A)**[28,29]. Consistent with *in vitro* measurements **(Fig. 1B),** we found no detectable uracil (Frac U) in HIV DNA isolated from resting CD4+ T cells**,** but five of the six donor monocyte samples contained detectable uracils. The fraction of uracil-positive *pol* amplicons was in the range 20 to 80% for the five patient samples **(Fig. 6B)**. The observation that viral DNA uracilation is specific to monocytes and macrophages provides an unambiguous marker of its origins and excludes contaminating T cell DNA as a possible source.

We also obtained access to cryopreserved bronchial alveolar lavage (BAL) and PBMC isolates from a single HIV-1 infected donor that were collected prior to initiation of ART and six months after ART therapy and performed a similar analysis (**Fig. 6B**). Viral DNA from highly purified monocytes was tested, as well as bronchial alveolar

lavage (BAL) samples that consisted of >95% alveolar macrophages (AMs). Using the Ex-Alu-*gag* PCR method for detecting proviral DNA we found that ~40% (monocytes) and 80% (alveolar macrophages) of the Alu-*gag* amplicons were positive for uracils. Similar levels of uracils were detected in pre- and post-ART BAL isolates (**Fig. 6B**). Although a stable reservoir for HIV is primarily located in resting CD4+ T cells, these results suggest that HIV DNA residing in monocytes/alveolar macrophages comprises another source for viral rebound after discontinued therapy, consistent with other reports[30].

**Discussion**

**Host cell DNA repair, HIV restriction, persistence and mutagenesis.** Previous literature contains diverse and sometimes conflicting data on the role of dUTP and host cell uracil DNA glycosylase in HIV infection [reviewed in Weil et al. [3]]. Our key finding that MDMs consist of two distinct populations with respect to the uracilation phenotype likely explains why this potent restriction pathway has remained largely elusive.

The established elements of this pathway are depicted in **Figure 6C**. Upon entry into the macrophage, reverse transcriptase (RT) encounters a nucleotide pool environment that favors the incorporation of dUTP into HIV DNA products predominantly in the form of U/A base pairs. The high ratio of dUTP/TTP is maintained, at least in part, by the low levels of dUTPase expression combined with the high expression levels of SAMHD1 in macrophages. Uracilation of the viral DNA can proceed to high levels in the cytoplasmic compartment because the UBER machinery is sequestered in the nucleus. However, when the heavily uracilated HIV DNA enters the

nucleus it is attacked and fragmented by nuclear UBER enzymes. Fragmentation is initiated by hUNG2 uracil excision, and is likely followed by abasic site excision by human abasic site endonuclease (APE1 or APE2) (**Fig. S1**) [31,32]. We estimate, based on comparing the measured levels of early and proviral DNA levels in the GFPMDMs (**Fig. 2**), that ~99% of the viral DNA is destroyed before the integration step. The viral DNA products that escape initial restriction can integrate into the host cell DNA and follow different fates that are discussed below (**Fig. 6C**).

We speculate that lethal mutagenesis, functional genome variation, or faithful repair could result from host enzyme processing of uracilated proviruses. In non-dividing MDMs, we found that uracilated proviruses contained few mutations yet gave rise to mutated viral genomic RNAs (**Fig. 4D**). Although it is tempting to attribute these mutations to RNAP II transcriptional errors as it encounters uracil on the template DNA strand, RNAP II shows high-fidelity incorporation of A opposite to U during *in vitro* transcription[25]. Moreover, many aspects of the mutation spectrum in **Fig. 4E** are not reconcilable with RNAP II incorporating incorrect bases opposite a template U (although the frequent A à G transitions may be explained by the potential of U to pair with A or G)[33]. The actual error mechanism is likely to be much more complex given that transcription coupled repair is expected to generate abasic sites, strand breaks, and even gaps when densely spaced uracils are excised. In this regard, RNAP II is known to stall at isolated abasic sites during *in vitro* transcription and could also misincorporate various ribonucleotides at such sites *in vivo*[34].

The observation that proviral DNA is relatively mutation free, while the RNA is not, strongly suggests that repair intermediates generated during transcription by the host UBER pathway are most often repaired to restore the proviral coding sequence. Faithful repair could involve the reincorporation of U opposite to A due to the perturbed nucleotide pools, or the introduction of T. The observation that RNA mutations were increased upon stimulation with the inflammatory cytokine IFNγ, and nearly absent in the presence of the anti-inflammatory cytokine IL4, indicates that the transcriptional programs induced by these cytokines are highly relevant to the outcome. Although errors introduced by IFNγ stimulation may frequently lead to non-functional viral genomes, they will also allow the viral DNA sequence to evolve rapidly in the absence of reverse transcription. Replication-competent viruses that emerge would then under go fitness selection in the host. These potential outcomes make UBER a double-edged sword with respect to propagation of viral infection.

Uracil has never been considered an epigenetic base like 5-substituted cytosines because of its low abundance in genomes and because the detection of U/A base pairs eludes standard sequencing methods. Nevertheless, uracil has the potential to alter DNA structure and dynamics [35], which could affect site recognition by transcription factors or other DNA binding proteins. There is a growing body of evidence indicating that uracil can exert significant effects on protein binding to DNA. These findings include (i) our previous report that multiple uracils silenced transcription from both HIV-1 LTR and CMV promoters in a cell line[3], (ii) uracils within the origin of replication in HSV-1 perturb the binding of HSV-1 origin binding protein[36], (iii) U/A pairs disrupt AP-1 transcription factor DNA binding[37], (iv) singly uracilated DNA disrupts RNase H splicing

specificity during reverse transcription [38], (v) U/A pairs perturb maintenance of telomere length in B cells by disruption of sheltrin binding [39], and (vi) one or two U/A base pairs within the specific cleavage site of some restriction enzymes prevents DNA strand cleavage[40].  In addition, the abasic site product of uracil excision is known to exert a large negative effect on transcription[41], as would any mutations in transcription factor recognition sequences arising from error prone repair of excised uracils [42,43]. Combined, these potential effects of U/A pairs could profoundly silence HIV gene expression.

Historically, macrophages have been controversial targets for HIV infection [44]. However, these long-lived cells have a significant impact on HIV-1 infection because they persist for months after initial infection [45], continue to produce infectious virus particles for their entire lifetime[46], and provide a drug tolerant environment for HIV propagation[47,48]). Moreover, their residence within tissues that have poor drug penetrance (lung, brain, gut)[30], has led to an increased interest in their potential role in the establishment and persistence of HIV infection.

One key question arising from our results is the lifetime of uracils within proviral DNA *in vivo*. The detection of abundant viral uracils in monocytes isolated from HIV-1 infected individuals who showed full viral suppression from ART, indicates that uracilation occurs *in vivo* during the normal course of viral infection of myeloid cells. The fairly short several day half-life of blood monocytes suggests that uracilated viral DNA within these cells arises from a recent infection, perhaps by passage of the monocyte through a tissue reservoir containing virus producing cells [28]. Such an infection pathway could involve a canonical CCR5-dependent viral entry mechanism or a recently

described pathway involving phagocytosis of infected T cells by macrophages[50]. However, the high efficiency of ART and the low levels of infected T cells using current treatment regimes make the phagocytosis pathway for monocyte infection seem less plausible for our aviremic patient samples. Finally, the requirement for dUTP incorporation during reverse transcription virtually excludes the possibility that the infected monocytes containing uracilated DNA arose from latently infected myeloid stem cells, because monocyte differentiation does not involve reverse transcription and genomic DNA replication does not involve accumulation of detectable uracils (**Fig. S8H**).

A further ramification of these data is that some uracilated proviruses may persist for the entire lifetime of a macrophage due to protection within highly compacted chromatin, which is highly resistant to UBER [35]. Under such circumstances, the potential time scale for uracil persistence could be years based on the findings that replication-competent viruses have been detected in brain microglial cells of patients even after years of viral suppression using ART [51,52]. If U/A pairs are found to down regulate viral gene expression through chromatin stabilization or by weakening the binding of transcription factors, some proviruses could remain transcriptionally silenced for years and then become activated when appropriate stimulatory conditions are present.

**Experimental Procedures**

**Study subjects**

This study was approved by the Johns Hopkins Institutional Review Board and written informed consent was provided by both HIV-infected and healthy individuals

(IRB00038590). Peripheral blood was obtained from healthy volunteers and ART suppressed patients with viral loads <20 copies HIV-1 RNA/mL. Monocytes and resting CD4+ T cells were obtained from donor PBMCs as described in the **Supplemental Methods** and the purity of the cell types was determined by flow cytometry.

**Viruses**

Pseudotyped HIV-1 virions (HIV$^{NL4.3}$ and HIV$^{SF162}$) were generated as previously described[3]. HIV$^{NL4.3}$(Δvpr) and HIV$^{NL4.3}$(Δvif) mutant virus constructs were generated by site-directed mutagenesis of pNL4-3-ΔE-eGFP. Primers used for site-directed mutagenesis can be found in **Supplemental Information (Table S3).**

**Western blot analyses**

Protein extracts were prepared from MDMs that had been mock treated or infected with HIV$^{NL4.3}$(wt) and HIV$^{SF162}$ or mutant viruses HIV$^{NL4.3}$(Δvpr) and HIV$^{NL4.3}$(Δvif). Protein extraction was performed using CelLytic M reagent (Sigma-Aldrich) according to the manufacturer's instructions. Protein concentration was determined by the Bradford assay (BioRad).

**Excision digital droplet PCR (Ex-ddPCR)**

To determine the uracil status of viral DNA a modified digital droplet PCR (Ex-ddPCR) method was developed by inserting a pre-digestion step using UNG (**Fig. 1**). Total

genomic DNA was isolated using a QIamp DNA extraction kit (Qiagen) according to manufacturer's protocol and processed as described in the **Supplemental Methods**.

**Ex-Alu-*gag* nested qPCR**

To specifically measure fractional uracilation of proviral DNA we used a modification of the Alu-*gag* nested qPCR approach, in which an UNG digestion step is inserted before the initial PCR amplification (Ex-Alu-*gag* nested qPCR)[3,16].

**Fluorescence activated cell sorting**

In preparation for cell sorting studies, MDMs infected with GFP-reporter virus (HIV[NL4.3] or HIV[SF162]) were washed twice with Hank's balanced salt solution (HBSS) and then detached by incubating with Accutase (Cell Technologies) for 30 min at 37 °C. MDMs were resuspended at $5\times10^6$/mL in 1x HBSS (pH 7.2), 5 mM EDTA and 0.5% BSA. Prior to sorting, cell suspensions were passed through a 35 mm nylon mesh (BD Biosciences) and propidium iodide added to gate out dead cells. Sorting was performed on a MoFlo Sorter using a 100 μm nozzle and the sort purify 1 mode.

**Viral RNA isolation and limiting dilution sequencing**

Viral RNA isolated from p24-positive wells was DNaseI-treated (Invitrogen) and reverse transcribed using the qScript cDNA synthesis kit (Quanta Biosciences). PCR amplifications were performed using limiting dilution conditions as previously described[21].

**Statistical Analyses**

Data were analyzed for statistical significance (HIV copy number and Frac U[DNA]) by a

two-tailed Student's t test for independent samples using GraphPad Prism. p<0.05 was deemed significant.

**Data Access**
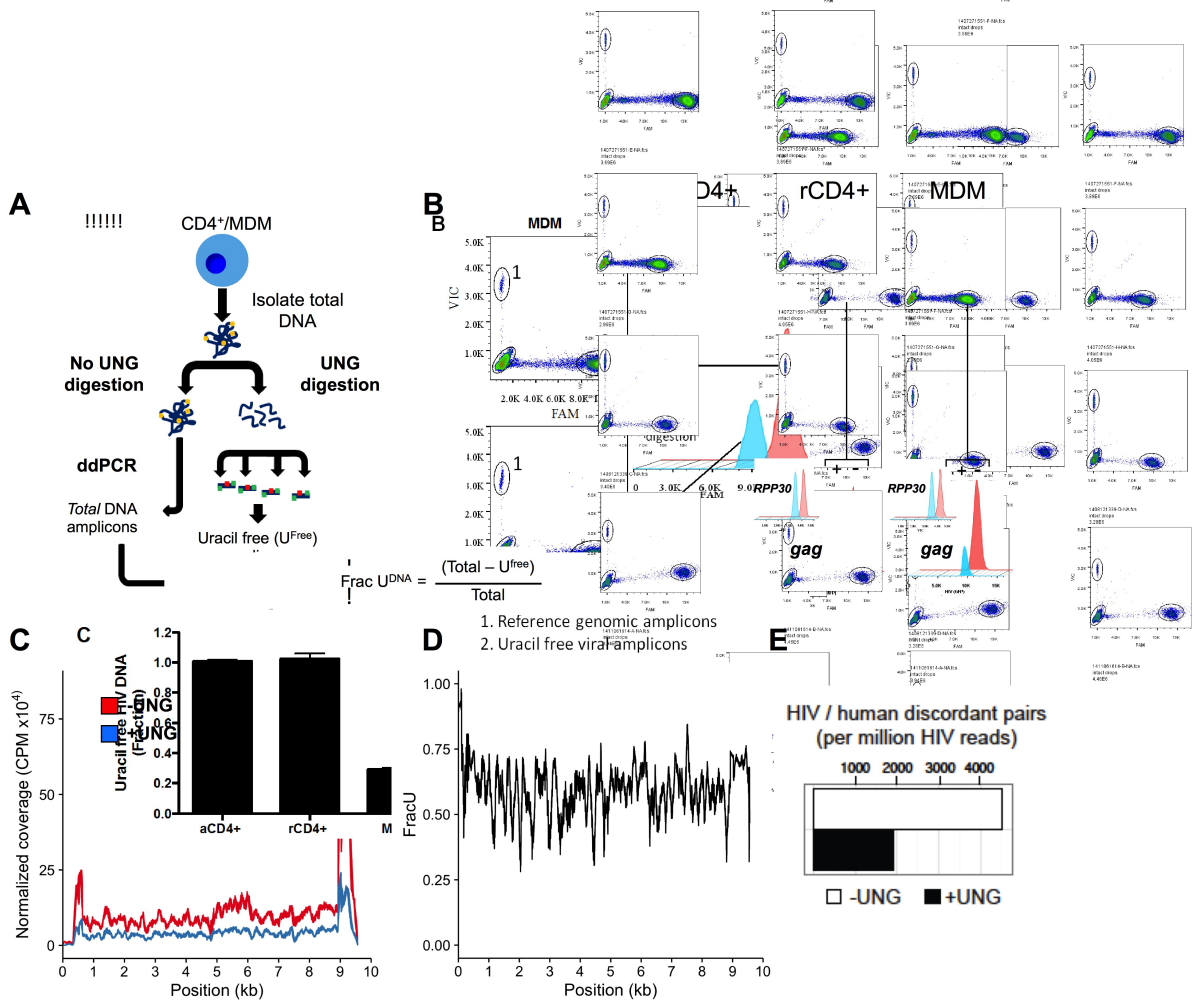
Raw and processed sequencing data files (FASTAQ) from this study have been deposited to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/)[53] under accession number GSE76091. A reproducible software pipeline for analysis of Excision-seq data is available at https://github.com/hesselberthlab/stivers-hiv.

**Figures**

**Figure 3.1 Ex-ddPCR to determine the uracil content of the HIV *gag* gene. (A)** Total DNA is isolated from immune cells and either digested by UNG to degrade uracil containing DNA or mock digested for total amplifiable DNA in the sample. The output (Frac U$^{DNA}$) represents the fraction of amplified DNA copies containing at least one uracil on each DNA strand. Signal is normalized to a genomic reference copy standard (*RPP30*) that does not contain uracil. **(B)** Activated, resting CD4+ T cells and MDMs were infected *in vitro* with HIV $^{NL4.3}$ virus and uracil content was measured 3 days post-infection (dpi) using primers that targeted the *gag* region. The data (+/- UNG digestion) are shown as scatter plots and histograms. **(C)** Normalized coverage of the HIV$^{NL4.3}$ genome positive strand in Excision-seq libraries prepared from total cellular DNA at 7 days post-HIV infection. **(D)** Fraction of the reads in panel C that contained uracil (Frac U). **(E)** Discordant read pairs between HIV and human DNA present in Ex-seq libraries prepared from total cellular DNA at 7 days post-HIV infection. The number of discordant reads obtained by Ex-seq in the absence and presence of UNG digestion are shown as white and black bars.
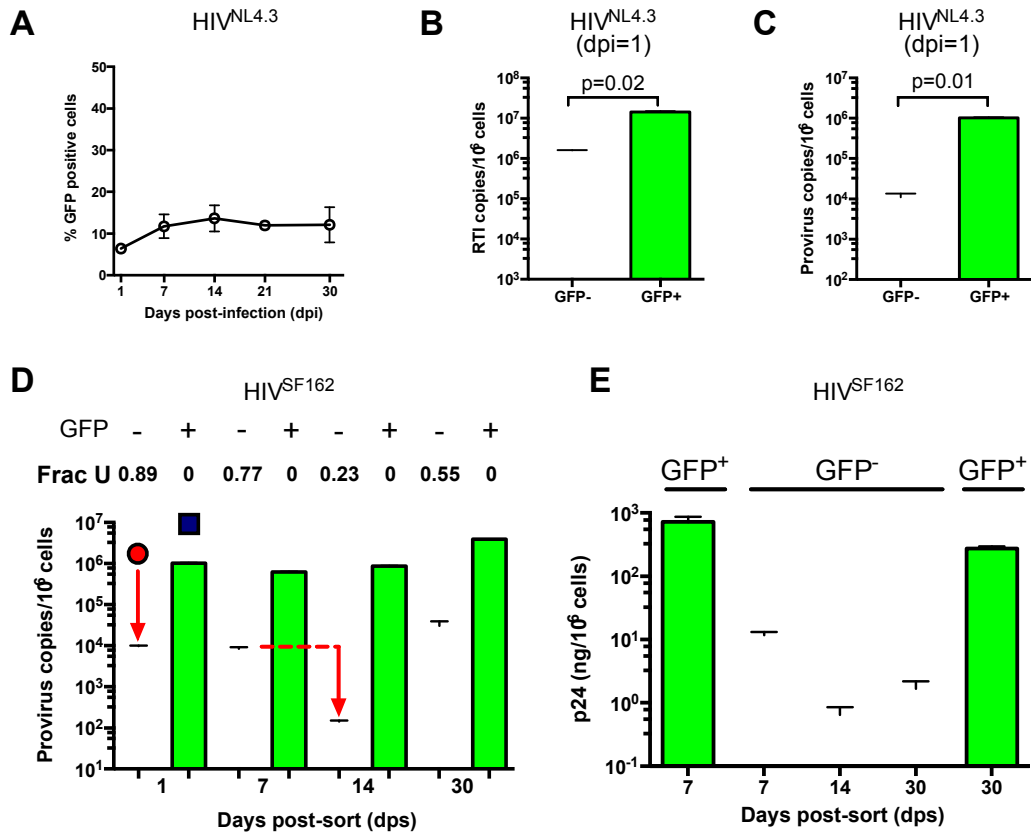
**Figure 3.2 MDMs consist of two distinct cell populations with respect to viral infection.** (**A**) Only about 10% of HIV[NL4.3] infected MDMs expressed virally encoded GFP as determined by flow cytometry. (**B**) ddPCR measurements of the copy number of reverse transcription intermediates (RTIs) for infected GFP⁻ (white bars) and GFP⁺ (green bars) MDMs. (**C**) Alu-*gag* nested qPCR measurement of the provirus copy number in GFP⁻ and GFP ⁺ MDMs. (**D**) Measurement of provirus copy number and uracil content over the course of a 30-day multi-round infection with HIV[SF162]. (**E**) ELISA measurements of viral p24 protein levels in GFP ⁻ and GFP+ MDMs over the course of a 30-day multi-round infection. Errors are reported as mean ± SD.
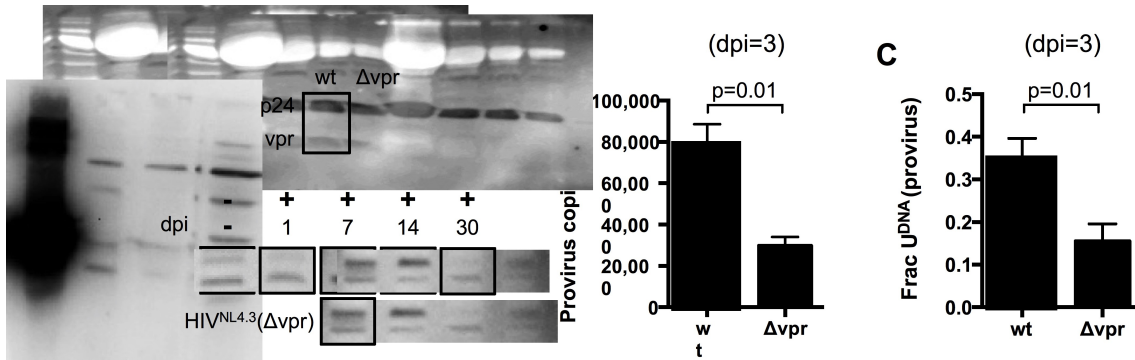
**hUNG2 uracil excision activity is antagonized by vpr.** (**A**) By Western blot the vpr deletion virus HIV$^{NL4.3}$(Δvpr) contained no detectable vpr. Western against the nuclear (hUNG2) and mitochondrial (hUNG1) isoforms was used to measure the selective oscillation in hUNG2 levels over the course of a 30 day multi-round infection. At day seven, the HIV$^{NL4.3}$(Δvpr) shows abundant hUNG levels while hUNG2 is depleted in the wild-type infection. (**B**) Measurements of provirus copy number in a mixed population of GFP$^-$ and GFP$^+$ MDMs infected with HIV$^{NL4.3}$ (wt) and HIV$^{NL4.3}$(Δvpr). (**C**) Ex-Alu-*gag* nested qPCR measurements of the fraction of uracilated proviral copies in a mixed population of GFP$^-$ and GFP$^+$ MDMs infected with HIV$^{NL4.3}$ (wt) and HIV$^{NL4.3}$(Δvpr) at three days post-infection. Errors are reported as mean ± SD.
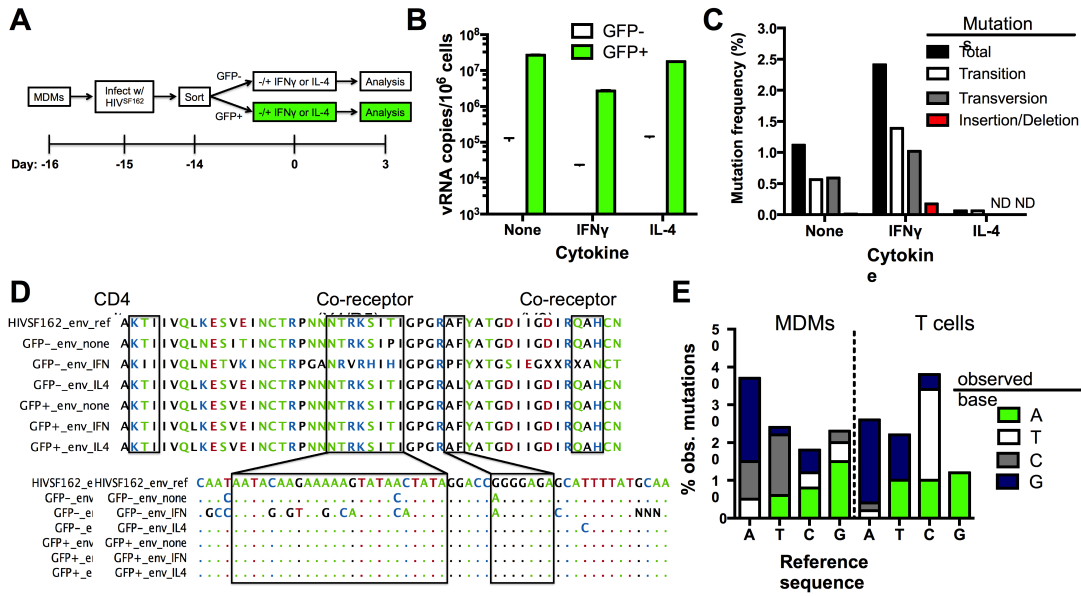
**Figure 3.4 Impact of uracilation and cytokines on viral genome sequence.**
(**A**) Experimental protocol. (**B**) Quantitative reverse transcriptase-PCR (qRT-PCR) was used to determine the copy numbers of extracellular viral RNA in cell supernatants from sorted GFP$^+$ (green bars) and GFP$^-$ (white bars) MDMs with and without cytokine stimulation. (**C**) Summary of mutations from limiting–dilution clonal sequencing of extracellular viral RNAs obtained from infected GFP MDMs under different growth conditions. The mutation frequencies (point mutations per total nucleotides sequenced) were obtained from a ~500 bp amplicon of the *env* gene. ND; not detected. (**D**) Representative *env* sequences of extracellular viral RNAs produced by GFP$^-$ sorted MDMs. The top sequence is for the HIV$^{SF162}$ strain used to infect the MDMs (HIVSF162_env_ref). Boxed regions at the protein and nucleotide sequence level show the mutation spectrum within the CD4-associated and co-receptor binding sites. (**E**) Breakdown of point mutations for viral RNA produced from infected MDMs and proviral DNA from infected T cells (see text).
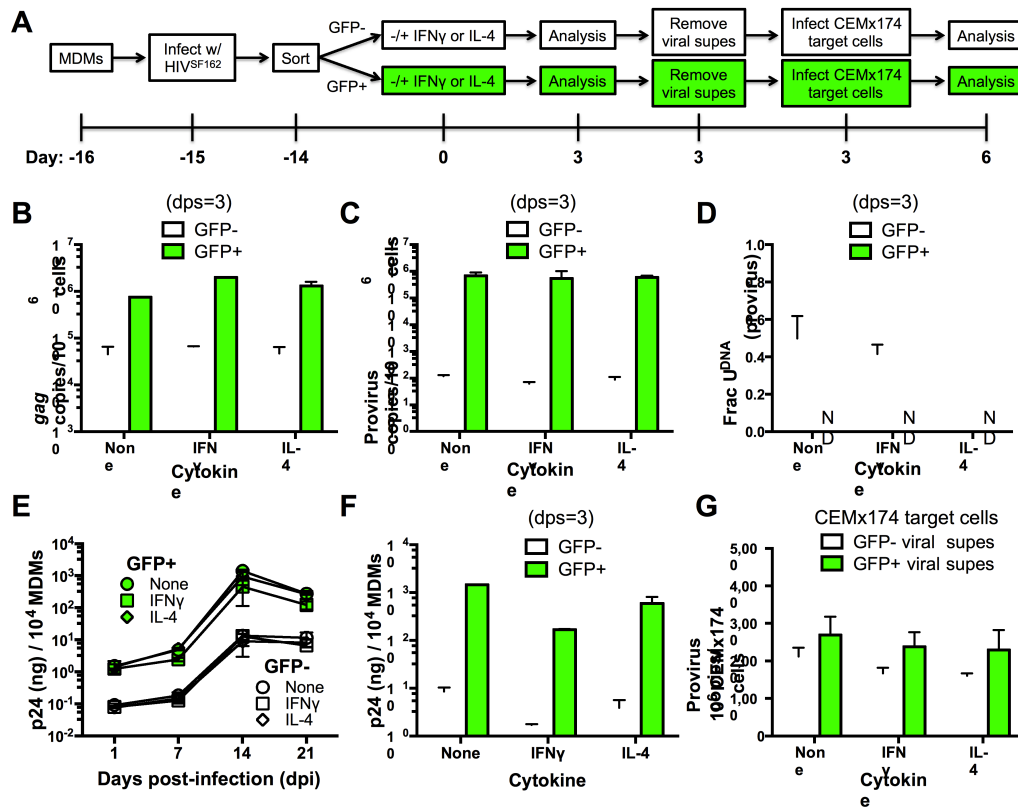
**Figure 3.5 Effects of cytokine stimulation on viral transmission in GFP sorted MDM populations.** (**A**) Experimental approach. (**B**) MDMs infected with HIV [SF162] were sorted into GFP+ (green bars) and GFP- (white bars) populations and analyzed with respect to *gag* copy number (dps; days post stimulation). (**C**) Provirus copy numbers for each MDM population were measured using Alu-*gag* nested qPCR at 3 dps. (**D**) The fraction of proviral DNA copies containing uracil was measured using Ex-Alu-*gag* nested qPCR. ND; not detected. (**E**) Viral growth kinetics of sorted MDM populations. (**F**) Levels of virus in culture supernatants of GFP+ and GFP- MDM producer cells were measured using p24 ELISA. (**G**) Viral supernatants from each MDM treatment were normalized to p24 levels and used to infect naïve CEMx174 target cells. Provirus copies were measured using Alu-*gag* qPCR. The results are normalized to one million target cells. Errors are shown as the mean ± SD.
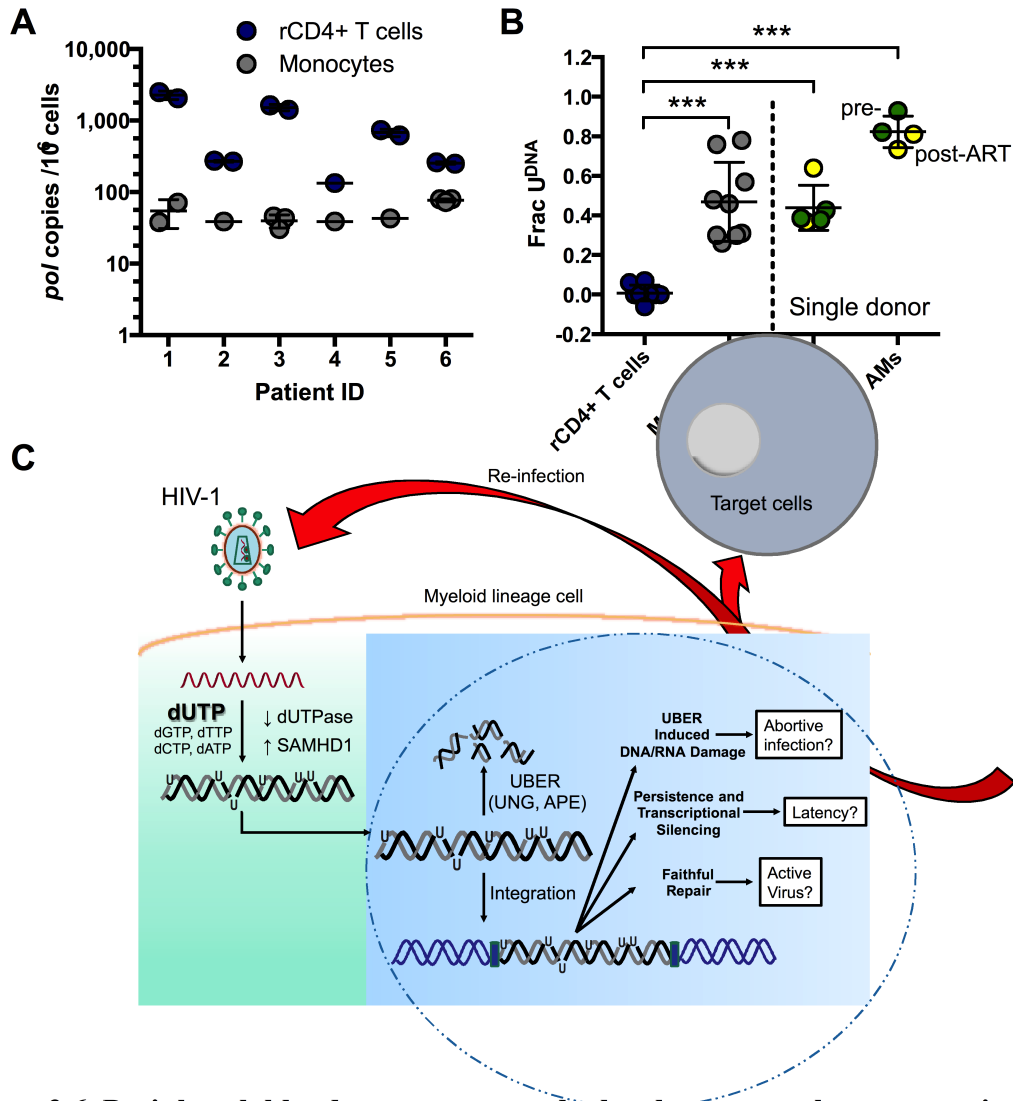
**Figure 3.6 Peripheral blood monocytes and alveolar macrophages contain high levels of uracil in HIV DNA. (A)** Resting CD4+ (rCD4+) T cells and monocytes were purified by negative bead selection from bulk PBMCs obtained from six ART-suppressed individuals. Total HIV*pol* DNA was quantified using Ex-ddPCR. Copy numbers were normalized to the human *RPP30* gene to give an estimate of HIV copies/10[6] cells. **(B)** The uracilated fraction of HIV *pol* DNA copies derived from monocytes and rCD4+ T cells was measured using Ex-ddPCR. Total genomic DNA was also isolated from matched, cryopreserved PBMCs and bronchial alveolar macrophages (AM) obtained from a single donor both pre-ART (green circle) and post-ART (yellow circle), (***p<0.001). Donor 4 had no detectable uracil (*pol*) and was excluded from this plot. **(C)** Infection of MDMs by HIV-1 and possible fates of uracilated viral DNA products (see text).
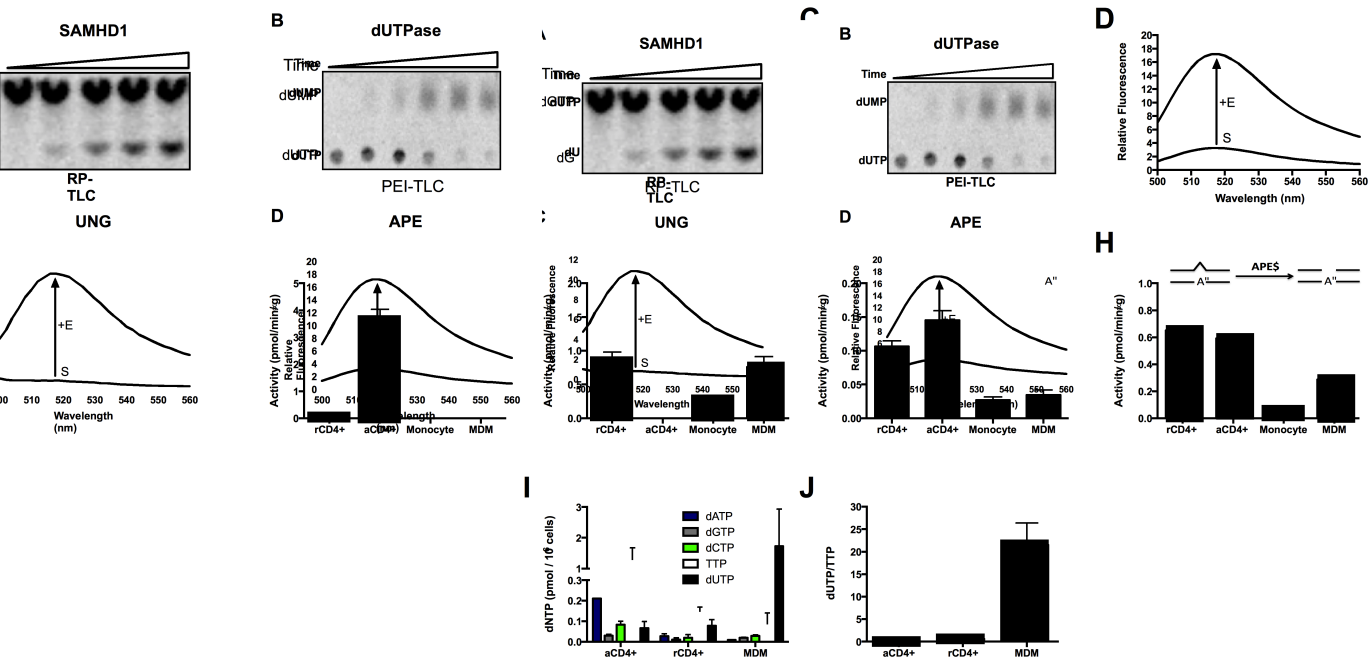
**Figure 3.7 Profiling enzyme activities and dNTP pool levels in immune target cells of HIV.** Extracts from each indicated cell type were obtained as described in Methods. (**A**) Deoxyuridinetriphosphate hydrolase (dUTPase) activity was measured by monitoring the hydrolysis of dUTP to dUMP via PEI-cellulose TLC. Specificity was determined using a potent dUTPase inhibitor (compound 26 in[16]). (**B**) SAMHD1 triphosphohydrolase activity was determined by C18 RP-TLC based assay using $^{3}$H-labeled dGTP as the substrate. Specificity for SAMHD1 was determined using the inhibitor pppCH$_2$dU. The mobilities of the substrate (dGTP) and product nucleoside (dG) are marked. (**C**) Endogenous uracil DNA glycosylase (hUNG) activity (combined hUNG1 and hUNG2) was determined using a fluorescein-labeled DNA substrate that shows an increase in fluorescence upon uracil excision. Specificity for hUNG activity was determined by addition of the uracil DNA glycosylase inhibitor protein (UGI). (**D**) Apyrimidinic endonuclease (APE1 or 2) activity was measured using a fluorescein-labeled duplex DNA containing a single abasic site that increases in fluorescence upon endonuclease cleavage. (**E**) dUTPase activity. (**F**) SAMHD1 activity. (**G**) hUNG activity. (**H**) APE activity. (**I**) Measurement of deoxyribonucleotide (dNTP) pool levels were determined by an LC-MS method. (**J**) MDMs contain a high ratio of dUTP/TTP ratio compared to resting and activated CD4+ T cells. Abbreviations: rCD4+, resting CD4+ T cell; aCD4+, PHA activated CD4+ T cell; MDM, monocyte-derived macrophage. See Supplemental Methods for further details and references describing these assays.
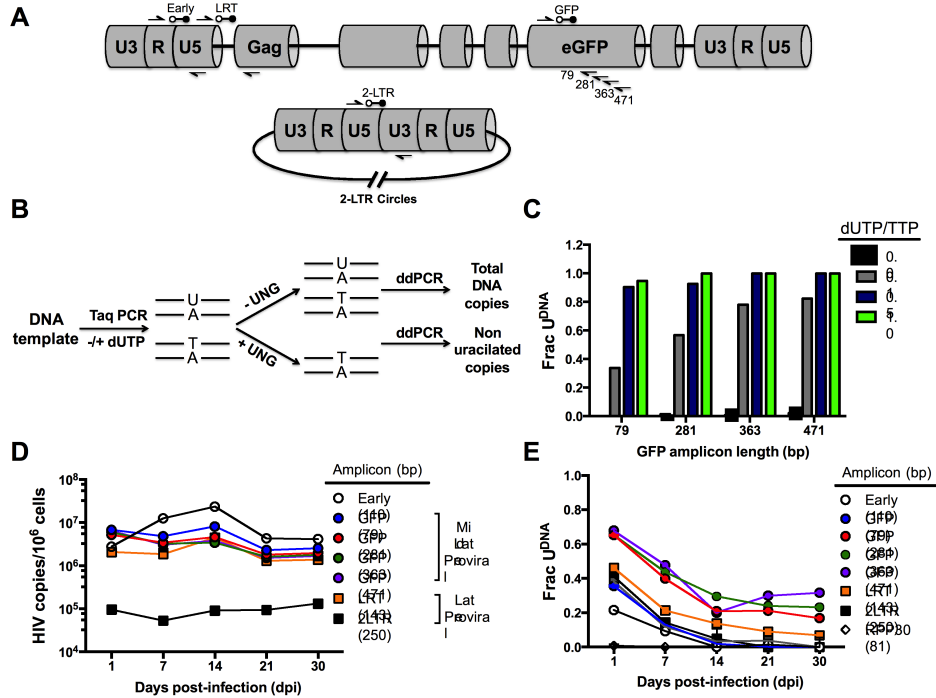
**Figure 3.8.** *In vitro* **generated calibration curves for evaluating uracil content in DNA amplicons and single-round HIV infections of cultured MDMs.** (**A**) ddPCR Primer and probe locations relative HIV genome (HIV[NL4.3]). (**B**) Generation of uracil containing duplex DNA of increasing length and ratio of dUTP/TTP was achieved by *in vitro* DNA polymerization using Taq polymerase in the presence of various ratios of dUTP/TTP. These DNA standards (79 to 471 bp) were analyzed by the Ex-ddPCR method as indicated. (**C**) The fraction of the DNA amplicons that contained detectable uracil (Frac U) increased with DNA amplicon length as well as the dUTP/TTP ratio used in the initial DNA synthesis. (**D**) In single-round infections of a mixed population of MDMs with HIV[NL4.3], the copy number of early, middle and late viral cDNAs were measured over a 30-day culture period using primers specific to different regions of the viral genome (see above). The copy number is normalized to one million MDM target cells. (**E**) Uracil content was measured in each HIV DNA population using Ex-Alu-*gag* nested qPCR.
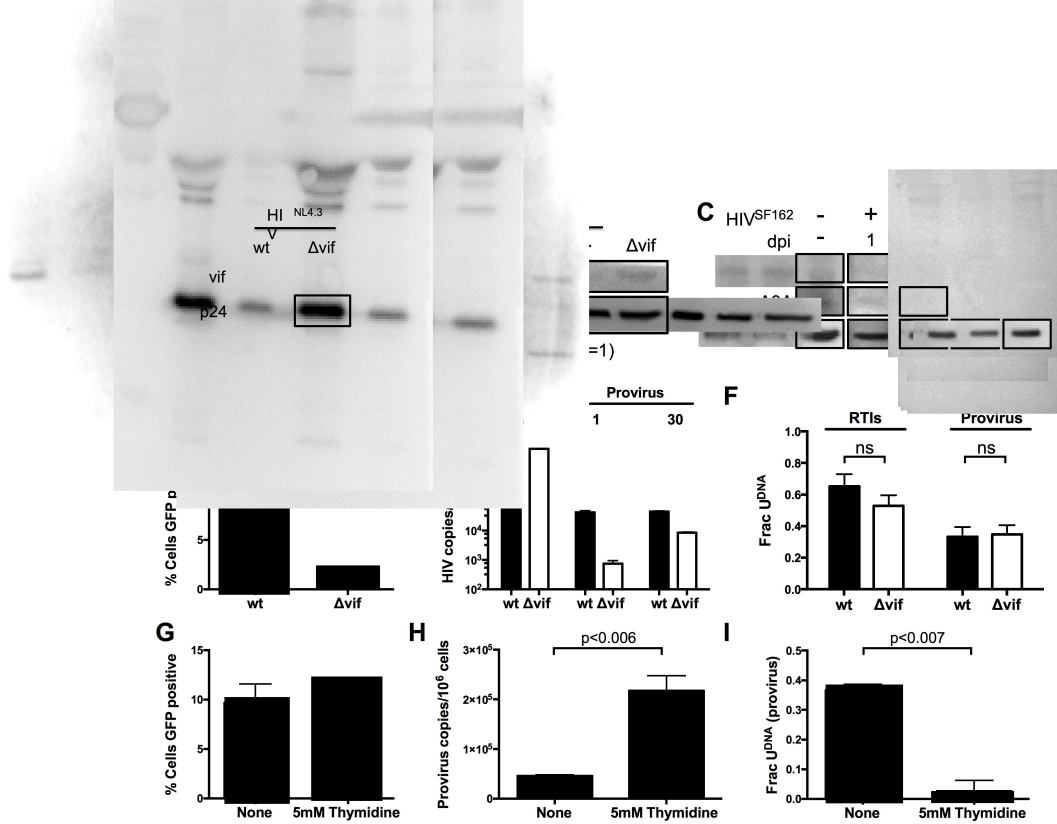
**Figure 3.** [...] **ination and can be rescued by thymidine supplementation. (A)** Western blots show no detectable vif in the deletion virus HIV$^{NL4.3}$(Δvif). **(B)** APOBEC3G (A3G) is poorly expressed in a mixed population of MDMs. Infection with wild type virus HIV$^{NL4.3}$ showed a small decrease in A3G, while infection with HIV$^{NL4.3}$(Δvif) did not affect the A3G levels. **(C)** Western blots established that the expression levels of APOBEC3A (A3A) and A3G in MDMs over a 30-day culture period after infection was consistently low. **(D)** Flow cytometry measurements established that viral GFP expression is decreased ~5-fold upon infection with HIV$^{NL4.3}$(Δvif) as compared to HIV$^{NL4.3}$. **(E)** Reverse transcription intermediates (RTIs) and provirus DNA levels are generally reduced for infection of MDMs with HIV$^{NL4.3}$(Δvif). Viral DNA copy numbers were measured by ddPCR (gag amplicon) and Alu-*gag* nested qPCR, respectively. **(F)** Ex-Alu-*gag* nested qPCR showed that the fraction of viral DNA copies containing uracil is unaffected when MDMs are infected with HIV$^{NL4.3}$(Δvif). **(G)** Supplementation of the culture media with 5 mM thymidine has no significant effect on GFP expression. **(H)** Cultures supplemented with 5 mM thymidine showed a 5-fold increase in provirus copy number as measured using Alu-*gag* nested qPCR. **(I)** Cultures supplemented with 5 mM thymidine result in a 14-fold decrease in the fraction of proviral copies that contain uracil as measured using Ex-Alu-*gag* nested qPCR.
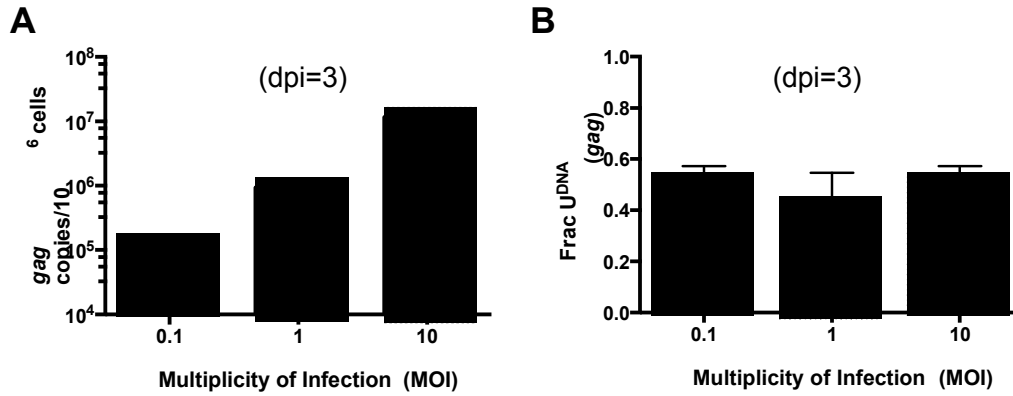
**Figure 3.10 Uracilation is independent of multiplicity of infection (MOI).** MDMs were infected with HIV$^{NL4.3}$ using MOIs of 0.1, 1 and 10. (**A**) The viral *gag* copy number increases with MOI. (**B**) The MOI does not significantly affect the fraction of viral gag copies that contained uracil.
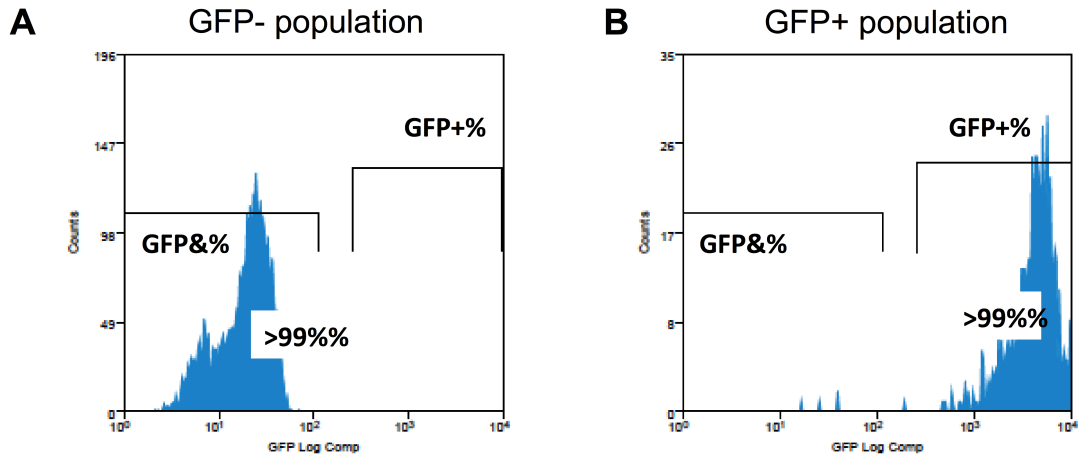
**Figure 3.11 The sorted populations of *in-vitro* infected GFP [-] and GFP [+] MDMs are highly pure.** MDMs were infected with HIV [NL4.3] by spinoculation and then sorted using GFP fluorescence after 7 days. The purity of each sorted population (GFP and GFP[+]) was assessed by flow cytometry. Each population was greater than 99% pure. (**A**) Flow cytometry analysis of purified GFP [-] population. (**B**) Flow cytometry analysis of purified GFP[+] population.
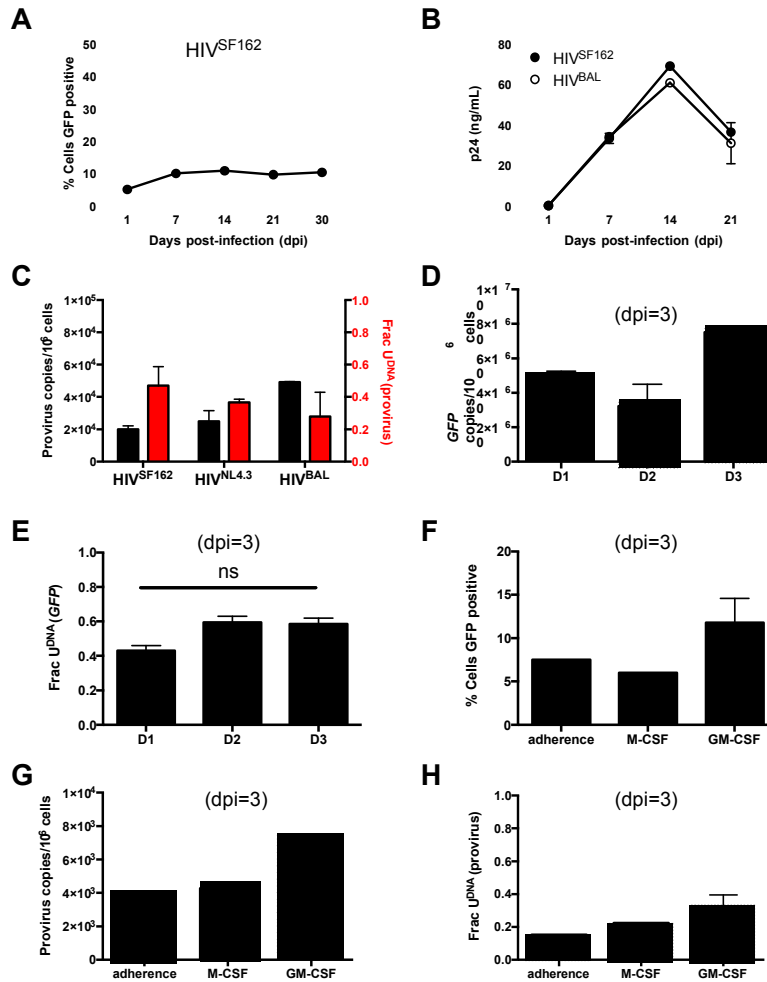
**Figure 3.12 Three different viral strains show a similar uracilation profile with *in vitro* infected MDMs independent of the differentiation regimen.** (**A**) Only a small fraction of infected MDMs (HIV$^{SF162}$) express virally encoded GFP as determined in flow cytometry. (**B**) The kinetics and absolute levels of virus output as measured by p24 ELISA is indistinguishable for MDMs infected with HIV $^{SF162}$ and HIV$^{BAL}$. (**C**) Provirus copy number and uracil content at dpi = 3 are very similar in MDMs infected with HIV$^{SF162}$, HIV$^{NL4.3}$ or HIV$^{BAL}$ (MOI = 0.1). Proviral copy numbers were measured using Alu-*gag* nested qPCR and the uracil content by Ex Alu-*gag* nested qPCR. (**D**) Infection of MDMs obtained from peripheral blood monocytes of three donors produces equivalent levels of uracilated HIV $^{NL4.3}$ DNA as judged by ddPCR analysis of the GFP copies. (**E**) The fraction of viral GFP copies that contain uracil is similar for all three donors as determined using Ex-ddPCR (ns; not significant). (**F**) Purified monocytes isolated from bulk PBMCs were differentiated to MDMs *in vitro* using the adherence method, or by addition of M-CSF or GM-CSF. The differentiation regimen did not significantly affect viral GFP expression in MDMs infected with HIV$^{SF162}$, or (**G**) proviral DNA copy number as determined by Alu-*gag* nested qPCR, or (**H**) the fraction of viral DNA copies that contained uracil as determined by Ex Alu-*gag* nested qPCR.
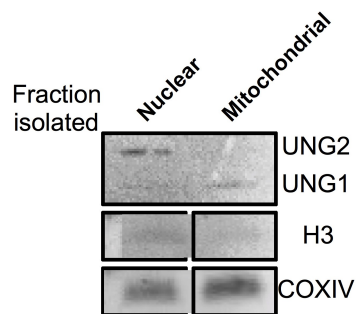
**Figure 3.13 Subcellular fractionation of nuclear and mitochondrial fractions from MDMs resolves the two UNG isoforms.** Western blot analysis detected both the mitochondrial (hUNG1) and nuclear (hUNG2) isoforms in MDMs. Cells were fractionated to isolate nuclear and mitochondrial fractions using the Qproteome kit (Qiagen). The purity of the isolated fractions was confirmed by Western probes of specific nuclear and mitochondrial protein markers, Histone H3 and COXIV, respectively.

**A** Monocytes

Isotype
CD14/CD16

CD16 / CD14

**B** Monocytes

Isotype
CD3

CD8

**C** Resting CD4+ T cells

Isotype
CD4

HLA-DR / FAM

VIC

**D**

| Patient ID | Isolated cell population | Purity (%) | HIV$^{pol}$ copies/10$^6$ cells |
|---|---|---|---|
| 1 | rCD4+ | 96 | 2268 ± 313 |
|  | Monocyte | 98.4 | 55 ± 23 |
| 2 | rCD4+ | 98 | 270 ± 6 |
|  | Monocyte | >99.9 | 39 |
| 3 | rCD4+ | 95 | 1519 ± 154 |
|  | Monocyte | >99.9 | 45 ± 2 |
| 4 | rCD4+ | 96 | 133 |
|  | Monocyte | >99.9 | 39 |
| 5 | rCD4+ | 95 | 676 ± 79 |
|  | Monocyte | >99.9 | 43 |
| 6 | rCD4+ |  | 387 ± 10 |
|  | Monocyte | >99.9 | 70 ± 13 |

| Sample Name | Name |
|---|---|
| 1511181528-F-NA.fcs | VIC |
| 1511181528-E-NA.fcs | VIC |
| 1511181528-D-NA.fcs | FAM |
| 1511181528-C-NA.fcs | FAM |

**E** Measured pol

Expected Count

**F** Measured RPP30

Expected RPP30 count

**G** Measured HIV

cells

**H** Frac UPN (RPP30)

○ Healthy donors
● ART patients

rCD4+ T cells    Monocytes

**RPP30**
☐ -UNG (orange)
☐ +UNG (green)

**pol**
☐ -UNG (red)
☐ +UNG (blue)

| Sample Name | Name |
|---|---|
| 1511181528-H-NA.fcs | VIC |
| 1511181528-G-NA.fcs | VIC |
| 1511181528-F-NA.fcs | FAM |
| 1511181528-E-NA.fcs | FAM |

VIC / FAM (Comp-PMT2)

VIC / FAM (Comp-PMT1)

| Sample Name | Name |
|---|---|
| 1511181528-H-NA.fcs | VIC |
| 1511181528-G-NA.fcs | VIC |
| 1511181528-F-NA.fcs | FAM |
| 1511181528-E-NA.fcs | FAM |

**Figure ... supp... d cell populations from ART** ... ls and monocytes were isolated from ... . The results for donor 1 are show ... ed for the specific monocyte mark ... preparations was determined by ne ... ermine if the purified resting CD4- ... ells, we stained for CD4 and the activation markers CD25/CD69/HLA-DR. (**D**) The purity of the purified monocyte and rCD4 cells from each patient is indicated. The HIV$^{pol}$ copies were determined by ddPCR and are expressed as an average and standard deviation (SD). When possible, two or three independent ... measurements were taken. (**E**) NL4.3 plasmid was serially diluted over ... valuate the detection limit of the ddPCR assay for HIV *pol*. (**F**) Dilution ... n genomic DNA with measurement of the *RPP30* copy number by dd ... s was used to determine the load limit for total DNA that can be accura ... a ddPCR reaction. (**G**) Measurement by ddPCR of HIV *pol* copy num ... und of genomic DNA obtained from 250,000 uninfected PBMCs. (**H**) ... t detected in genomic DNA (*RPP30* gene) of uninfected monocytes o ... ng RPP30 primers, Ex-ddPCR was performed on genomic DNA isolated from monocytes of healthy donors (open circles) or ART patients (gray circle) as well as rCD4+ T cells from ART patients. (**I**) Ex-ddPCR was used to determine that uracil is only found in HIV DNA isolated from monocytes and not rCD4+ T cells of HIV-infected patients. The shown data are for donor ID 2 (**Table S4**). For visualization, the histograms for the RPP30 Ex-ddPCR data have been down scaled by a factor of ~10$^4$ to facilitate comparison with the low abundance *pol* positive droplets. The insets show the raw scatter plots of the *pol* positive droplets that were used to generate corresponding

histograms for each isolated cell population. The *pol* positive droplets for the ddPCR and Ex-ddPCR experiments are shown in red and blue, respectively. The histograms for the *pol* positive droplets follow the same color key.

o

o

o

o

o

o

o

o

o

o

**Supplemental Methods**

**Ex-vivo enzyme activity assays**

Protein lysates were obtained using CelLytic M (Sigma) reagent according to the manufacturer's instructions. SAMHD1 activity was determined as previously described with several modifications[1]. Protein lysates (5 μg) were incubated in buffer containing; 5 mM [8-$^3$H] dGTP (activator/substrate) (Moravek Biochemicals), 5 mM ATP, and 10 mM paranitrophenyl phosphate (p-NPP) in a total volume of 50 μL. ATP was included to inhibit nonspecific phosphatases and p-NPP was added to prevent the degradation of dUTP by alkaline phosphatase[2]. The small molecule inhibitor (pppCH$_2$dU) of SAMHD1 was used to determine specificity of the assay[3]. Reactions were incubated at 37°C at varying time points, at which two microliter samples were removed and quenched by spotting onto a C18-reversed phase thin layer chromatography (TLC) plate. The TLC plate was developed in 50 mM KH$_2$PO$_4$ (pH 4.0) to separate substrate [dGTP ($R_f$ = 0.80)] from products [dG ($R_f$ = 0.20)]. Plates were exposed to a tritium sensitive screen for 5 h and then scanned on a Typhoon phosphoimager (GE Healthcare) and the counts present in the substrate and product were quantified using the program Quantity One (Bio-Rad).

dUTPase activity was assessed as previously described[4]. Briefly, protein lysates were incubated with [5-$^3$H] dUTP (Moravek Biochemicals) and the fraction of substrate hydrolyzed to dUMP product was monitored. Reactions were incubated at 37°C for 1 h, after which two microliter samples was spotted onto a PEI-cellulose TLC plate. The TLC

plates were developed in 0.5 M LiCl to separate substrate [dUTP $_f(R_f$ 0.1)] from the product [dUMP ($R_f$ = 0.6)]. Plates were exposed to a tritium sensitive screen for 5 h and scanned on a Typhoon phosphoimager (GE Healthcare).

A molecular beacon hairpin reporter assay was used to determine endogenous UNG activity in crude cell extracts[4]. Reactions were performed using 5 µg of protein lysate and 50 nM S-pin-18 in 10 mM Tris-HCl (pH 7.1), 100 mM NaCl, 1 mM EDTA, and 0.2% Triton X-100. Supplementing the reaction with the potent Uracil DNA glycosylase inhibitor (UGI) showed no measurable activity. Additionally, a molecular beacon assay was used to monitor Apurinic/Apyrimidinic endonuclease (APE) activity [5]. Two FAM and dabcyl quenched DNA duplex substrates were used. A specific substrate (SS) containing a single abasic site (Φ) and a non-specific substrate (NS) with the abasic site replaced with the canonical DNA base Thymine (T). The sequence of the specific molecular beacon substrate; SS, 5'-FAM-GAGAAΦATAGTCGCG3' and 5'-CGCGACTATGTTCTC-dabsyl-3' (where Φ is a tetrahydrofuran abasic site analog). Independent reactions were performed using 5 µg of protein lysate and 50 nM in 10 mM Tris-HCl (pH 7.1), 100 mM NaCl, 1 mM EDTA, and 0.2% Triton X-100. To determine APE activity, the difference in the initial rate of the specific substrate from the non-specific rate was measured. Rates were measured on a Fluoro-Max 3 (Horiba Jobin Yvon).

**dNTP Extraction, rNTP removal and LC-MS Quantification**

The quantification of dNTPs from primary immune cells was performed by quantitative

LC-MS. To remove the signal depression by the abundant rNTPs [6], a boronic acid chromatography step was performed prior to analysis. Briefly, NTPs were methanol extracted for 1 hr at 4°C from 5 million cells. Samples were supplemented with isotopically labeled $^{13}C^{15}N$-dNTPs (Sigma-Aldrich) for accurate determination of each metabolite. Supernatants were dried under vacuum and then resuspended in 50 µL (200 mM Ammonium acetate, pH 8.8). This was directly applied to 100 µL of pelleted Boronic Acid resin (0.5 mg/mL) and incubated at 4°C for 1 hr. Resin was pelleted by centrifugation, supernatant removed and dried under vacuum. Sample was resuspended in mobile phase A [2 mM $NH_4PO_4H_2$, 3 mM Hexylamine (HMA) in dd $H_2O$]. Chromatographic separation was performed with an analytical Hypersil Gold C18 column (100 x 1 mm, 3 µm particle size, Thermo Scientific), using a (Agilent) LC-MS system equipped with a binary pump. Mobile phase A was [2 mM $NH_4PO_4H_2$, 3 mM Hexylamine (HMA) in dd $H_2O$] and mobile phase B [Acetonitrile]. The flow rate was maintained at 50 ∝L/min and an injection volume of 20 ∝L. The autosampler was held constant at 4°C and the column at 30°C. The chromatographic gradient began at 10 % mobile phase, followed by a linear gradient that reached 60 % in 15 min. At the end of each run, the column was equilibrated for 10 min. Analyte detection was performed with the same LC-MS system with an electrospray ionization source, using multiple-reaction monitoring (MRM) analysis in positive ionization mode. The following MRM transitions (parent à product) were detected as previously described with the addition of dUTP: dATP (492 *m/z* à 136 *m/z*), dGTP (508 *m/z* à 152 *m/z*), dCTP (468 *m/z* à 112 *m/z*), TTP (483 *m/z* à 81 *m/z*) and dUTP (469 *m/z* à 81 *m/z*)[7]. A standard curve was constructed using equimolar amounts of dNTPs.

*In vitro* **infection of primary immune cells**

Infections with HIV viral strains (NL4.3, SF162, BAL) were performed as previously described[4]. Briefly, primary cells were plated at a density of ~100,000 cells per well in a 96-well plate. Virus was added to each well and the plate was spin infected for 2 hrs at 1,200 x g and 30°C and then incubated at 37°C until the given time points. Following incubation, cells were collected and stored at -80°C.

**Western blots for detection of APOBEC3A and APOBEC3G**

Detection of both APOBEC3A and APOBEC3G was achieved by preparing crude lysates as described in *Western blots for vpr and vif*. Twenty-five micrograms of each sample was run on an SDS/PAGE gel and transferred to a PVDF membrane. Primary antibody (1:1000) was used for detection of APOBEC3G and is cross-reactive with APOBEC3A. Membrane was incubated overnight with anti-APOBEC3G C-terminal 29 amino acid obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (cat. no. 10201; from Immunodiagnostics: Murine Human APOBEC3G (CEM15) Monoclonal Antibody. The secondary HRP-conjugated anti-rabbit IgG (1:2000) was used for detection. Global protein expression was assessed by blotting MDM extracts for GAPDH.

**Western blots for detection of *vif* and *vpr* in mutant viruses**

The primary antibody for GAPDH (1:2000) used was a rabbit polyclonal antibody (ab15246; Abcam), and the secondary antibody (1:5000) was HRP-conjugated anti-rabbit IgG (ab97080; Abcam). Ten micrograms of each sample was run on an SDS/PAGE gel and transferred to a PVDF membrane. Detection of virion-associated viral protein R (vpr)

was achieved using a rabbit polyclonal antibody (1:2000) obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (cat. no.11836; from Dr. Jeffrey Kopp) and viral infectivity factor (vif) using mouse monoclonal antibody (1:2000) obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (cat. no. 319; from Dr. Michael H. Malim) the secondary antibody (1:5000) was HRP-conjugated goat anti-mouse IgG (ab97040; Abcam). Virion associated proteins were probed by re-suspending purified virus directly in buffer containing SDS. For an internal control, the same membrane was also blotted for p24 using the mouse monoclonal anti-p24 (1:2000) primary antibody obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (cat. no. 24-2; from Dr. Michael H. Malim). The secondary antibody (1:5000) used was an HRP-conjugated goat anti-mouse IgG (ab97040; Abcam).

**Subcellular fractionation of MDM cells**

Isolation of mitochondrial and nuclear fractions was achieved using the Qproteome Mitochondria Isolation kit (37502; Qiagen) according to the manufacturer's protocol. The rabbit polyclonal antibody against the C-terminus of hUNG (62520; Abcam) was used to detect both mitochondrial (hUNG1) and nuclear (hUNG2) isoforms. To ensure equal loading and transfer, membranes were probed for the mitochondrial specific protein cytochrome oxidase subunit IV (COXIV) (9715S; Cell Signaling) and nuclear Histone 3 (H3) (4850S; Cell Signaling). Immunoreactive proteins were visualized using HRP linked goat anti-mouse antibody (ab97040; Abcam).

**Ex-ddPCR method and determination of the fraction of uracil in HIV DNA**

Genomic DNA was first fragmented using the endonuclease BSAJ-1 (1 U) in Cutsmart

buffer (NEB) for 1 hr at 60°C. Uracil containing DNA sequences were digested using 50

nM UNG in 1x TMNB+ buffer (10 mM Tris-HCl (pH 8.0) 20 mM NaCl, 11 mM $MgCl_2$

and 0.002% Brij-35). Reactions were cleaned up using a MinElute PCR Purification kit

(Qiagen). UNG or mock digested DNA were diluted in to a PCR master mix containing

(1x Platinum taq buffer, 1x droplet stabilizer (RainDance), 1 mM dNTPs, 0.9 μM

forward and reverse primers, 0.15 μM probe, 2 mM $MgCl_2$, 0.5 μM carboxy-X-

rhodamine (ROX, Sigma) as a passive reference dye, 1 U platinum taq polymerase and

varying amounts of template DNA) and loaded in to a RainDance source chip

(RainDance Technologies). Different primer/probe sets were designed to tile the HIV

genome to report on heterogeneity of uracil incorporation. Amplification was performed

with the following thermal program: 95°C for 10 min, followed by 44 cycles of: 95°C for

15 sec and then 60°C for 1 min, followed by a final step of heating to 95°C for 10 min.

The final high-temperature cycle cures the droplets. For each step, the ramp cycle time

was decreased to 0.5 °C/s. Amplification samples were transferred to a RainDance Sense

instrument. Copy numbers of the human RNase P (*RPP30*) gene were measured in the

same reaction mixture to determine cell number. DNA isolated from the Jurkat-based cell

line (J-Lat)[8] containing a full-length integrated HIV genome was used to establish gating

parameters for HIV and RPP30 positive droplets. J-Lat cells were obtained through the

NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: J-Lat Full Length Clone

(clone 9846) from Dr. Eric Verdin. Ex-ddPCR data were analyzed using the RainDrop

Analyst software package (RainDance Technologies). Drop size gating was performed to

remove drops of atypical size using a lower and upper size gate boundary for each PMT

of 800 and 1150, respectively. Intact drops for each of the three clusters (one negative;

quenched drops, and two positive; PMT1+ and PMT2+) were gated independently using

the ellipse-gating tool. Gated clusters were spectrally compensated using the following

definitions: (intact drops/quenched drops) as the negative control for both PMT1+ and

PMT2+, while (intact drops/PMT1+) and (intact drops/PMT2+) were selected for PMT1+

and PMT2+, respectively. Positive cluster gating on spectrally compensated data was

again achieved by using the ellipse-gating tool. A Poisson correction statistic was applied

to both positive targets (HIV and *RPP30*) to remove sampling error [9]. Determination of

uracil free DNA transcripts was determined using Eq 1.

$$ FracU^{!"\#} \quad = \frac{\{!"\#\$\%\$\&'()"!*'\%\# !("\#\$\% !!"\#)\$\%\$\&'()"!*'\%\# !!"\# \quad ( \qquad )\}}{\{!"\#\$\%\$\&'()"!*'\%\# !("\#\$\% \quad )\}} \tag{1} $$

**Excision sequencing (Ex-seq) library preparation**

Ex-seq libraries were constructed from genomic DNA by shearing for 12 minutes on

BioRuptor (Diagenode), end repair, dA-tailing, and adapter ligation (NEB) with dual

indexed adapters (GSE76091). Library concentrations were quantitated with a Qubit

fluorometer and mixed at equal proportions to yield 500 ng of library DNA. HIV

enrichment was performed using lockdown probes designed against the forward and

reverse strand of the HIV construct used for infection, HIV[NL4.3] (GSE76091). Briefly,

libraries were mixed with 5 µg Cot-1 and XGen Blocking oligos and hybridized with the

lockdown probes at 3 pmol total for 4 hours at 65°C. Probes were isolated on

Streptavidin Dynabeads (M-270; Invitrogen) to capture HIV DNA using Nimblegen

SeqCap EZ Hybridization and Wash kit as indicated in the XGen lockdown probe protocol. DNA was eluted in water and samples were split in half, followed by uracil removal in one sample with UNG (5 U) and T4 Endonuclease IV (10 U) digestion for 2 hours at 37°C on beads. On-bead PCR was performed for 22 cycles with short Illumina primers (GSE76091) and Maxima Taq DNA polymerase (Invitrogen). PCR products were purified with 1.9x volumes of Ampure XP beads (Agencourt) and quantified with a Qubit fluorometer. Forward and reverse lockdown reactions were pooled for UNG- and UNG+ independently and two lanes of Illumina HiSeq were performed. Data files from this study have been deposited to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession GSE76091.

**Excision-seq data analysis**

Demultiplexed FASTQ records were trimmed to remove low quality cycles (cycles 1 and 2, and 70-125 for HT29 samples), and processed records were aligned to the human (hg19) and HIV reference (HIV $^{NL4.3}$) with bwa mem (unpublished). Alignments in BAM format were converted to bedGraph coverage with BEDTools. Samples were deduplicated with samtools rmdup and Piccard markduplicates. VCF files were generated for mutation analysis from the rmdup files with freebayes and filtered with vcffilter for quality scores greater than 40[10]. Mutations were further filtered for a depth of greater than 10 reads and 10 or more unique sequencing start sites. Chimeric and discordant reads were identified with samtools and were normalized to counts per million reads that mapped to HIV. A reproducible software pipeline for analysis of Excision-seq data is available at https://github.com/hesselberthlab/stivers-hiv.

**Ex-Alu-*gag* nested PCR for uracil detection**

Ex-Alu-*gag* nested PCR begins with the first amplification (40 cycles) used a primer complementary to genomic Alu sequences and a *gag* primer and ~100 ng genomic DNA. In addition to samples, reactions containing: *gag* primer only, no-template, and genomic DNA isolated from uninfected matched cells were used as controls. A standard curve for Ex-Alu-*gag* nested qPCR was generated using DNA isolated from J-Lat cells. Although there is some debate concerning the measurement of absolute proviral copy numbers using Alu-*gag* nested qPCR, Ex-Alu-*gag* nested qPCR is unambiguous with respect to determining the change in copy number resulting from UNG digestion. This is due to the fact that identical DNA samples are used for the PCR amplification steps. Thus the fraction of the proviral population that contains uracil within the *gag* amplicon may be reliably measured.

**Cytokine stimulation of cultured MDMs**

FACS sorted (GFP-/+) MDM populations were maintained in culture media for indicated times prior to cytokine stimulation. MDMs were either stimulated by co-culture with the IFNγ (50 ng/mL, R&D Systems) and 200 ng HIV $_{tat}$ peptide (ProSpec Ltd.) or IL-4 (50 ng/mL, R&D Systems). At three days post-stimulation (dps), the cytokine containing media is removed and replaced with fresh culture media. Viral supernatants were collected at varying time points from each well and assayed for HIV-1 p24 production using a sensitive ELISA (Alliance HIV-1 p24 antigen ELISA kit, Perkin Elmer).

**Isolation of primary cells from donor PBMCs**

Peripheral blood mononuclear cells (PBMCs) were isolated from HIV-1 positive or negative donors using density centrifugation on a Ficoll-Hypaque gradient. Monocytes were isolated from the PBMC population using the Pan monocyte Isolation Kit (Miltenyi Biotec). Monocytes were differentiated into macrophages over 7 days using MDM-20 media containing RPMI 1640, 20% (vol/vol) autologous plasma, 10 ng/mL GM-CSF or M-CSF (BD Biosciences), 1× HEPES, and 1× Glutamine (Gibco). Cultured MDMs were maintained in media containing RPMI 1640 + 10% (vol/vol) dialyzed FBS, 1% Pen/Strep. CD4+ T cells were isolated from PBMC population using the CD4+ isolation kit (CD4+ T cell Isolation kit II, Miltenyi Biotec). CD4+ cells were activated with 0.5 µg/mL phytohemagglutinin (PHA) for 3 days in IL-2 containing media. Resting CD4+ cells were further enriched from the bulk CD4+ cell population by negative selection against biotinylated activation markers; CD25, CD69 and HLA-DR (Miltenyi Biotec).

**Purity of isolated cell populations**

The purity of monocytes, activated and resting CD4+ cells was determined by flow cytometry with fluorescent labeling of anti-human; Fluorescein isothiocyanate (FITC)- and allophycocyanin (APC)-conjugated mAbs CD4, CD25, CD69 and HLA-DR (BD Biosciences). Isotype-matched control mAbs (BD Biosciences) were used for gating and quantification. Monocyte and MDM populations were stained using brilliant violet 421-conjugated (BV421) anti-CD14, CD16-APC and CD3-FITC. The purity of these cells were analyzed by flow cytometry using a FACS Canto II (BD Biosciences) and FlowJo software (Treestar). Using this method we obtained monocytes with a purity of >99.9% with undetectable T cell contamination. Cells were then pelleted and used for protein extraction, dNTP extraction, or infection as described above; for cell infections, cells

were analyzed for GFP expression by FACS analysis.

**_In vitro_ infection and nested limiting dilution DNA sequencing of CD4+ T cells**

T cells were infected with replication-competent HIV [BAL] as previously described above. The HIV[BAL] (R5) virus was obtained through NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH (cat. no. 510; contributors: Dr. Suzanne Gartner, Dr. Mikulas Popovic and Dr. Robert Gallo). Briefly, CD4+ T cells isolated from PBMCs of an HIV-1 negative donor were purified using a negative selection method (CD4[+] T cell Isolation Kit II, Miltenyi Biotec) and activated with PHA for 3 days in IL-2-containing medium. A spinoculation method (2 hrs, 37°C, 1200 x g) was used to infect activated CD4+ T cells with replication-competent HIV[BaL] virus (300 ng p24/$10^6$ cells). Following incubation for 6 days in IL-2 containing media, supernatants were collected and stored at -80°C. Freshly isolated CD4+ T cells were activated by αCD3/αCD28 stimulation for 3 days in IL-2 containing medium and then plated (100 µL/well) into a 96-well v-bottom plate. HIV [BaL] (500 ng p24/$10^6$ cells) was added to each well and cells were infected by spinoculation. Following infection, cells were suspended in 40mL of IL-2 containing medium supplemented with enfuvirtide (10 µM) to prevent additional rounds of viral replication. Following a two day incubation at 37°C, genomic DNA was extracted from $10^7$ CD4[+] cells using the Qiagen Gentra Purgene Cell Kit A. DNA was subjected to a nested limiting dilution PCR protocol adapted from _Ho et al_. with the following modifications: 30 cycles were run for the outer PCR, 40 cycles were run for each inner PCR, and all clonal outer wells were subjected to the 4 inner PCRs (A-D; corresponding to gag, pol, rev and env regions) regardless if positive or negative for the gag inner PCR[11].

**Extracellular viral RNA limiting dilution sequencing**

HIV infected (HIV$^{SF162}$) MDMs were sorted into GFP$^-$ and GFP$^+$ populations and diluted serially in 5-fold increments ($1x10^5$ – 100 cells) using 12-well plates (Costar). Cells were maintained in culture for 14 days prior to cytokine simulation and supernatant collection. Extracellular viral RNA was collected from the supernatants of individual wells that contained the fewest number of input cells and were p24-positive in an attempt to obtain virus released from a cell containing a single provirus. Extracellular viral RNA was purified using a ZR-96 Viral RNA Kit (Zymo Research). Isolated viral RNAs were DNase-treated (Life Technologies) and reverse transcribed using a qScript cDNA synthesis kit (Quanta Biosciences) according to the manufacturer's protocol. Using the products of RT-PCR, nested PCR reactions were used to amplify the LTR and *env* regions of interest (primer sequences are listed in **Table S3**). Limiting dilution PCR was performed using Platinum Taq High-Fidelity polymerase (Invitrogen). For LTR amplification, outer primers 5LTROut and mod_VQA_R and inner primers 5LTRIn and mod_VQA_R were used. Thermocycler settings were the same for both outer and inner LTR PCR reactions: 94°C for 2 min, followed by 29 cycles of: 94°C for 15 sec, 55°C for 30 sec and then 68°C for 1 min, followed by a final extension at 68°C for 7 min. For *env* amplification, outer primers ES7 and ES8 and inner primers Nesty8 and DLoop were used. Thermocycler settings were the same for both outer and inner *env* PCR reactions: 94°C for 3 min, followed by 25 cycles of: 94°C for 30 sec, 55°C for 30 sec and then 68°C for 1 min, followed by a final extension at 68°C for 5min.

cDNA was diluted (1/20 – 1/1000) and used as input for the outer PCR reaction. Aliquots (1μL) from each outer PCR product were used as input for inner PCR reactions

and subjected to 1% agarose gel electrophoresis. Inner clonal PCR reactions from selected dilutions were used to identify dilutions where <20% of the PCR reactions were positive, where the corresponding outer PCR dilution contained one template with >90% of possibility by Poisson statistics. PCR products were gel extracted using the QIAquick Gel Extraction Kit (Qiagen) and directly sequenced (Sanger) without cloning (Genewiz). Forward and reverse sequences for each sample were aligned into one consensus contig per sample using default assembly parameters with CodonCode Sequence assembly and Alignment software and aligned to $HIV^{SF162}$ reference sequence. Chromatograms showing double peaks were taken as evidence for more than one template present in the initial PCR reaction and were discarded from further analysis.

**Bronchio-alveolar lavage (BAL) cell processing**

Bronchoscopy and broncho-alveolar lavage was performed on a single HIV-1 infected patient pre and post-ART treatment. At the time of the first bronchoscopy the patient was ART naïve with recent plasma viral load 24,162 copies/ml and CD4 459 cells/mm$^3$. Second bronchoscopy occurred 9.2 months later, after 6 months of ART treatment (efaviernz/emtricitabine/tenofovir) with plasma viral load <20 copies/ml and CD4 413 cells/mm$^3$. To obtain alveolar macrophages (AMs), bronchoscopy and lavage were done, as described elsewhere[12]. Briefly, lavage fluid was collected, filtered and BAL cells were pelleted by centrifugation. BAL cells were washed with HBSS and counted. After centrifugation, cells were re-suspended in Recovery Cell Culture freezing media according to the BAL count (10$^6$ cells/mL) and stored in a -140°C freezer. PBMCs were

isolated on a ficoll-hypaque gradient and were processed in the same manner as BAL cells.

**Determination of HIV *pol* copies and uracilation in *in vivo* infected monocytes and resting T cells using ddPCR and Ex-ddPCR**

Monocytes and resting CD4+ T cells purified via negative selection (Miltenyi Biotec) from blood collected from six HIV infected patients on antiretroviral therapy (ART). Cell purity was determined as described above in *Purity of isolated cell populations*. Cellular DNA was extracted using a QIamp DNA Midi Kit (Qiagen) following the manufacturer's protocol. The frequency of total HIV DNA (*pol* copies/$10^6$ cells) was determined by ddPCR using published primers to conserved regions of HIV *pol* (Althaus, 2010) and the reference cellular gene RNase P (*RPP30*) for genomic quantification. Total cellular DNA was estimated by halving the number of *RPP30* copies and HIV copy numbers per diploid cell were calculated as the ratio of template *pol* copies per diploid cell (*pol* copies/$10^6$ cells). Due to the low frequency of HIV DNA in a large background of host cellular DNA it was desirable to load as much total DNA (~$10^6$ cells) as possible to maximize assay sensitivity. See **Fig. S10** for further details.

The limit of detection and quantification of the ddPCR assay was evaluated by establishing the loading limit and intrinsic limit of detection. A dilution series of uninfected PBMC DNA was used to determine the maximum amount of total DNA that could be accurately measured in a single ddPCR reaction. Cellular *RPP30* was used to monitor input DNA and showed a linear response across a wide range ($10^3$-$10^6$) of *RPP30* copies. The intrinsic limit of detection of (5 copies/$10^6$ cells) was determined by serially diluting DNA from infected CD4+ T cells into a background of healthy donor

122

PBMC DNA (250,000 cells). For patient samples, replicate measurements were taken to improve the accuracy of the measurement.

**Supplemental Data**

**Uracils do not arise from APOBEC cytosine deamination**

It has been previously reported that MDMs do not express high levels of APOBEC3G (A3G) or APOBEC3A (A3A) DNA cytidine deaminases, and that the limited levels present reside in the inactive high-molecular weight (HMW) form [13,14]. Nevertheless, we wanted to completely exclude the possibility that the detected uracils arose significantly from APOBEC cytosine deamination.

Since viral vif protein is known to target A3G for proteasomal degradation, we reasoned that the viral DNA products produced from a vif deletion virus should have

increased uracilation levels due to increased APOBEC activity. Although vif serves other functions that promote infection of MDMs [15], which would be expected to result in lower overall levels of infectivity upon its deletion, these effects would not change the uracil content of viral DNA products. Accordingly, we infected MDMs with wild-type HIV$^{NL4.3}$ virus or the vif deletion mutant, HIV$^{NL4.3}$($\Delta$vif), which contains no detectable packaged vif protein (**Fig. S3A**). Infection of MDMs with these viruses produced only small changes in the basal level of A3G in cell extracts, possibly due to the resistance of HMW-A3G to vif (**Fig. S3B**). Similarly, minor decreases in the already low basal expression levels of A3G and A3A were observed in cell extracts obtained before and after infection with the replication competent HIV$^{SF162}$ virus (**Fig. S3C**).

In contrast with the small effects on A3G and A3A levels, HIV$^{NL4.3}$($\Delta$vif) showed a 4-fold decrease in GFP expression as compared to wild-type virus at one day post-infection (**Fig. S3D**). Although infections with the $\Delta$vif virus produced the same levels of early DNA products as wild-type (**Fig. S3E**), proviral DNA appeared at a much lower level as judged by Alu-*gag* nested qPCR (**Fig. S3E**). Importantly, no increases in the uracilation level of early viral DNA products or proviral DNA were observed upon infection with HIV$^{NL4.3}$($\Delta$vif) (**Fig. S3F**).

If the viral uracils in MDMs arose from dUTP incorporation, we anticipated that the addition of thymidine to the culture media prior to infection would lead to an increase in the viral copy number (less uracil related pre-integration restriction) and less uracil present in proviral DNA due to increased competition by TTP during reverse transcription (**Fig. S3G, 3H, 3I**). Although the presence of 5 mM thymidine produced no significant change in GFP expression (**Fig. 3G**), it increased the proviral copy number by

5-fold (**Fig. 3H**), and decreased the uracilation level of proviral DNA by 14-fold (**Fig. 3I**). The combined results strongly support the conclusion that the viral uracils arise primarily from dUTP incorporation and are therefore in the form of coding U/A base pairs.

**Tables**

**Table S1.** Characteristics of GFP$^-$ and GFP$^+$ MDM populations[a]

| Sorted population | % Total cells | Frac U$^{DNA,b}$ | hUNG[c] | dUTPase[c] | SAMHD1[c] | dUTP/TTP |
|---|---|---|---|---|---|---|
| GFP$^-$ | ~90 | ~90% | $0.06 \pm .01$ | N/D | $0.7 \pm 0.1$ | 20/1 |
| GFP$^+$ | ~10 | N/D[d] | $0.08 \pm .02$ | N/D | $0.8 \pm 0.2$ | 1.1/1 |

[a]Infected MDMs were sorted into two populations according to GFP fluorescence at 1 day post-infection (HIV$^{SF162}$). [b]Fraction of proviral DNA copies that contained uracil by Alu-*gag* nested qPCR. Enzymatic activities are presented as pmol/min/μg. [d]N/D = Not detected.

**Table 1.1 Characteristics of GFP$^-$ and GFP$^+$ MDM populations**

**Table S2.** Mutational analysis of proviral DNA isolated from MDMs [a] and CD4+ T cells[b]

| Sorted population | Region | Cytokine | Clones sequenced | Base pairs sequenced | Total point mutations | Transition mutations | Transversion mutations | I+D[c] |
|---|---|---|---|---|---|---|---|---|
| GFP[-] | LTR | None | 20 | 6,160 | 3 | 2 | 2 | 0 |
| | | IFNγ | 27 | 8,316 | 2 | 0 | 0 | 0 |
| | | IL-4 | 32 | 9,856 | 2 | 0 | 0 | 0 |
| | *env* | None | 8 | 3,264 | 0 | 0 | 0 | 0 |
| | | IFNγ | 14 | 5,712 | 0 | 0 | 0 | 0 |
| | | IL-4 | 14 | 5,712 | 0 | 0 | 0 | 0 |
| GFP[+] | LTR | None | 23 | 7,084 | 0 | 0 | 0 | 0 |
| | | IFNγ | 21 | 6,468 | 0 | 0 | 0 | 0 |
| | | IL-4 | 15 | 4,620 | 0 | 0 | 0 | 0 |
| | *env* | None | 19 | 7,752 | 0 | 0 | 0 | 0 |
| | | IFNγ | 17 | 6,936 | 0 | 0 | 0 | 0 |
| | | IL-4 | 16 | 6,528 | 0 | 0 | 0 | 0 |
| T cells | LTR | None | 14 | 4312 | 39 | 5 | 2 | 0 |
| | *env* | None | 14 | 5712 | 50 | 5 | 1 | 0 |

[a]Infected MDMs containing proviruses were sorted into two populations according to GFP fluorescence at 14 days post infection and then activated with the indicated cytokines. Mutation frequencies are calculated based on total nucleotides sequenced (i.e. frequency = mutated nucleotide count/total nucleotides). [b]*In* vitro infected resting CD4+ T cells. [c]Insertion and deletions counted on a per event basis.

**Table 1.2 Mutational analysis of proviral DNA isolated from MDMs and CD4+ T cells**

**Table S2 (cont.)** Mutational analysis of extracellular viral RNA produced from GFP⁻ and GFP⁺ MDMs[a]

| Sorted population | Region | Cytokine | Clones sequenced | Base pairs sequenced | Total point mutations | Transitions mutations | Transversion mutations | I+D |
|---|---|---|---|---|---|---|---|---|
| GFP⁻ | LTR | None[d] | 5 | 1,540 | 0 | 0 | 0 | 0 |
| | | IFNγ | 5 | 1,540 | 0 | 0 | 0 | 0 |
| | | IL-4 | 14 | 4,312 | 0 | 0 | 0 | 0 |
| | env | None[c] | 10 | 4,080 | 48 | 7 | 0 | 3 |
| | | IFNγ | 6 | 2,448 | 56 | 9 | 3 | 7 |
| | | IL-4 | 8 | 3,264 | 2 | 0 | 0 | 0 |
| GFP⁺ | LTR | None | 7 | 2,156 | 0 | 0 | 0 | 0 |
| | | IFNγ | 7 | 2,156 | 0 | 0 | 0 | 0 |
| | | IL-4 | 7 | 2,156 | 0 | 0 | 0 | 0 |
| | env | None | 11 | 4,488 | 0 | 0 | 0 | 0 |
| | | IFNγ | 9 | 3,672 | 0 | 0 | 0 | 0 |
| | | IL-4 | 9 | 3,672 | 0 | 0 | 0 | 0 |

[a]Infected MDMs containing proviruses were sorted into two populations according to GFP fluorescence at 14 days post infection and then activated with the indicated cytokines. Mutation frequencies are calculated based on total nucleotides sequenced (i.e. frequency = mutated nucleotide count/total nucleotides). RT-PCR was used to amplify extracellular viral RNAs into cDNA prior to clonal sequencing.

**Table 1.2 (Cont.) Mutational analysis of proviral DNA isolated from MDMs and CD4+ T cells**

**Table S3**. Primer and molecular beacon probe sequences (5' à 3')

| | |
|---|---|
| GFP forward | CGACTTCTTCAAGTCCGCCA |
| GFP reverse (79bp) | CTTGTAGTTGCCGTCGTCCT |
| GFP reverse (281bp) | TGCCGTCCTCGATGTTGTG |
| GFP reverse (363bp) | CTCAGGTAGTGGTTGTCGGG |
| GFP reverse (471bp) | TTGTACAGCTCGTCCATGCC |
| GFP probe | FAM-CCCGAAGGCTACGTCCAGGAGCGC-BHQ2 |
| Early forward | GCTAACTAGGGAACCCACTGCTT |
| Early reverse | CAACAGACGGGCACACACTGCTT |
| Early probe | FAM-AGCCTCAATAAAGCTTGCCTTGAGTGCTTC-BHQ2 |
| Alu forward | GCCTCCCAAAGTGCTGGGATTACAG |
| gag forward (6F) | CATGTTTTCAGCATTATCAGAAGGA |
| gag reverse (84R) | TGCTTGATGTCCCCCCACT |
| gag probe | FAM-CCACCCCACAAGATTTAAACACCATGCTAA-BHQ2 |
| LRT forward (MH531) | TGTGTGCCCGTCTGTTGTGT |
| LRT reverse (MH532) | GAGTCCTGCGTCGAGAGATC |
| LRT probe | FAM-CAGTGGCGCCCGAACAGGGA-BHQ2 |
| 2LTR forward (MH535) | AACTAGGGAACCCACTGCTTAAG |
| 2LTR reverse (MH536) | TCCACAGATCAAGGATATCTTGTC |
| 2LTR probe (MH603) | FAM-ACACTACTTGAAGCACTCAAGGCAAGCTTT-BHQ2 |
| pol forward (mf299) | GCACTTTAAATTTTCCCATTAGTCCTA |
| pol reverse (mf302) | CAAATTTCTACTAATGCTTTTATTTTTTC |
| pol probe (mf348) | FAM-AAGCCAGGAATGGATGGCC-MGBNFQ |
| RPP30 forward | GATTTGGACCTGCGAGCG |
| RPP30 reverse | GCGGCTGTCTCCACAAGT |
| RPP30 probe | VIC-CTGACCTGAAGGCTCT-MGBNFQ |
| vpr sense | GGAAACTGACAGAGGACAGGTGGAACAAGCCCC |
| 5LTRout | CACACAAGGCTAYTTCCCTGA |
| Mod VQA R | TTTTTTGAGGCTTAAGCAGTGGGTTCCCTA |
| ES7 | CTGTTAAATGGCAGTCTAGC |
| ES8 | CACTTCTCCAATTGTCCCTCA |
| Nesty8 | CATACATTGCTTTTCCTACT |
| DLoop | GTCTAGCAGAAGAAGAGG |
| vpr antisense | GGGGCTTGTTCCACCTGTCCTCTGTCAGTTTCC |
| vif sense | GGAAACTGACAGAGGACAGGTGGAACAAGCCCC |
| vpr antisense | GGGGCTTGTTCCACCTGTCCTCTGTCAGTTTCC |

**Table 1.3 Primer and molecular beacon probe sequences (5' à 3')**

**References**

1.  Priet, S., Sire, J. & Querat, G. Uracils as a Cellular Weapon Against Viruses and Mechanisms of Viral Escape. *Curr HIV Res* **4,** 31–42 (2006).
2.  Sire, J., Querat, G., Esnault, C. & Priet, S. Uracil within DNA: an actor of antiviral immunity. *Retrovirology* **5,** 45–12 (2008).
3.  Weil, A. F. *et al.* Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc Natl Acad Sci USA* **110,** E448–E457 (2013).

4.  Yan, N., O'Day, E., Wheeler, L. A., Engelman, A. & Lieberman, J. HIV DNA is heavily uracilated, which protects it from autointegration. *Proc Natl Acad Sci USA* **108,** 9244–9249 (2011).

5.  Kennedy, E. M. *et al.* Abundant Non-canonical dUTP Found in Primary Human Macrophages Drives Its Frequent Incorporation by HIV-1 Reverse Transcriptase. *J Biol Chem* **286,** 25047–25055 (2011).

6.  Bregeon, D., Doddridge, Z. A., You, H. J., Weiss, B. & Doetsch, P. W. Transcriptional Mutagenesis Induced by Uracil and 8-Oxoguanine in Escherichia coli. *Mol Cell* **12,** 959–970 (2003).

7.  Krokan, H. E. & Bjoras, M. Base Excision Repair. *Cold Spring Harb Perspect Biol* **5,** a012583–a012583 (2013).

8.  Krokan, H. E. *et al.* Error-free versus mutagenic processing of genomic uracil— Relevance to cancer. *DNA Repair* **19,** 38–47 (2014).

9.  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 USA* **111,** E1843–E1851 (2014).

10. Seiple, L. A., Cardellina, J. H., Akee, R. & Stivers, J. T. Potent inhibition of human apurinic/apyrimidinic endonuclease 1 by arylstibonic acids. *Mol Pharmacol* **73,** 669–677 (2008).

11. Goldstone, D. C. *et al.* HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480,** 379–382 (2011).

12. Gavegnano, C., Kennedy, E. M., Kim, B. & Schinazi, R. F. The Impact of Macrophage Nucleotide Pools on HIV-1 Reverse Transcription, Viral Replication, and the Development of Novel Antiviral Agents. *Mol Biol Int* **2012,** 1–8 (2012).

13. Hollenbaugh, J. A. *et al.* dNTP pool modulation dynamics by SAMHD1 protein in monocyte-derived macrophages. *Retrovirology* **11,** 1–12 (2014).

14. Bryan, D. S., Ransom, M., Adane, B., York, K. & Hesselberth, J. R. High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res 24,* 1534-1542

15. Sabbatucci, M. Endogenous CCL2 neutralization restricts HIV-1 replication in primary human macrophages by inhibiting viral DNA accumulation. *Retrovirology* **12,** 1–22 (2015).

16. O'Doherty, U., Swiggard, W. J., Jeyakumar, D., McGain, D. & Malim, M. H. A Sensitive, Quantitative Assay for Human Immunodeficiency Virus Type 1 Integration. *J Virol* **76,** 10942–10950 (2002).

17. Eldin, P. *et al.* Vpr expression abolishes the capacity of HIV-1 infected cells to repair uracilated DNA. *Nucleic Acids Res* **42,** 1698–1710 (2014).

18. Grogan, B. C., Parker, J. B., Guminski, A. F. & Stivers, J. T. Effect of the Thymidylate Synthase Inhibitors on dUTP and TTP Pool Levels and the Activities of DNA Repair Glycosylases on Uracil and 5-Fluorouracil in DNA. *Biochemistry* **50,** 618–627 (2011).

19. Xue, J. *et al.* Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. *Immunity* **40,** 274–288 (2014).

20. Laird, G. M. *et al.* Rapid Quantification of the Latent Reservoir for HIV-1 Using a Viral Outgrowth Assay. *PLoS Pathog* **9,** e1003398 (2013).

21. Ho, Y.-C. *et al.* Replication-Competent Noninduced Proviruses in the Latent

Reservoir Increase Barrier to HIV-1 Cure. *Cell* **155,** 540–551 (2013).

22. Rose, P. P. & Korber, B. T. Detecting hypermutations in viral sequences with an emphasis on G -> A hypermutation. *Bioinformatics* **16,** 400–401 (2000).

23. Yu, Q. *et al.* Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11,** 435–442 (2004).

24. Hildinger, M. *et al.* Membrane-Anchored Peptide Inhibits Human Immunodeficiency Virus Entry. *J Virol* **75,** 3038–3042 (2001).

25. Menéndez-Arias, L. Mutation Rates and Intrinsic Fidelity of Retroviral Reverse Transcriptases. *Viruses* **1,** 1137–1165 (2009).

26. Cassetta, L. *et al.* M1 polarization of human monocyte-derived macrophages restricts pre and postintegration steps of HIV-1 replication. *AIDS* **27,** 1847–1856 (2013).

27. Strain, M. C. *et al.* Highly Precise Measurement of HIV DNA by Droplet Digital PCR. *PLoS ONE* **8,** e55943–8 (2013).

28. Wang, T. *et al.* Successful Isolation of Infectious and High Titer Human Monocyte-Derived HIV-1 from Two Subjects with Discontinued Therapy. *PLoS ONE* **8,** e65071–12 (2013).

29. Eriksson, S. *et al.* Comparative Analysis of Measures of Viral Reservoirs in HIV-1 Eradication Studies. *PLoS Pathog* **9,** e1003174–17 (2013).

30. Cribbs, S. K., Lennox, J., Caliendo, A. M., Brown, L. A. & Guidot, D. M. Healthy HIV-1-Infected Individuals on Highly Active Antiretroviral Therapy Harbor HIV-1 in Their Alveolar Macrophages. *AIDS Res. Hum. Retroviruses* **31,** 64–70 (2015).

31. Schrader, C. E., Guikema, J. E. J., Wu, X. & Stavnezer, J. The roles of APE1, APE2, DNA polymerase and mismatch repair in creating S region DNA breaks during antibody class switch. *Philos Trans R Soc B Biol Sci* **364,** 645–652 (2009).

32. Stavnezer, J. *et al.* Differential expression of APE1 and APE2 in germinal centers promotes error-prone repair and A:T mutations during somatic hypermutation. *Proc Natl Acad Sci USA* **111,** 9217–9222 (2014).

33. Kuraoka, I. Effects of Endogenous DNA Base Lesions on Transcription Elongation by Mammalian RNA Polymerase II. Implications for transcription-coupled DNA repair and transcriptional mutagenesis. *J Biol Chem* **278,** 7294–7299 (2002).

34. Yu, S. L., Lee, S. K., Johnson, R. E., Prakash, L. & Prakash, S. The Stalling of Transcription at Abasic Sites Is Highly Mutagenic. *Mol Cell Biol* **23,** 382–388 (2003).

35. Ye, Y. *et al.* Enzymatic Excision of Uracil Residues in Nucleosomes Depends on the Local DNA Structure and Dynamics. *Biochemistry* **51,** 6028–6038 (2012).

36. Focher, F., Verri, A., Verzeletti, S., Mazzarello, P. & Spadari, S. Uracil in Oris of Herpes simplex 1 alters its specific recognitionby origin binding protein (OBP): does virus induced uracil-DNA glycosylase play a key role in viral reactivation and replication? *Chromosoma* **102,** S67–S71 (1992).

37. Rogstad, D. K., Liu, P., Burdzy, A., Lin, S. S. & Sowers, L. C. Endogenous DNA Lesions Can Inhibit the Binding of the AP-1 (c-Jun) Transcription Factor. *Biochemistry* **41,** 8093–8102 (2002).

38. Klarmann, G. J., Chen, X., North, T. W. & Preston, B. D. Incorporation of uracil into minus strand DNA affects the specificity of plus strand synthesis initiation

during lentiviral reverse transcription. *J Biol Chem* **278,** 7902–7909 (2003).

39.    Vallabhaneni, H. *et al.* Defective repair of uracil causes telomere defects in mouse hematopoietic cells. *J Biol Chem* **290,** 5502–5511 (2015).

40.    Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* **43,** D298–9 (2015).

41.    Luhnsdorf, B., Epe, B. & Khobta, A. Excision of Uracil from Transcribed DNA Negatively Affects Gene Expression. *J Biol Chem* **289,** 22008–22018 (2014).

42.    Shah, S., Pirrone, V., Alexaki, A., Nonnemacher, M. R. & Wigdahl, B. Impact of Viral Activators and Epigenetic Regulators on HIV-1 LTRs Containing Naturally Occurring Single Nucleotide Polymorphisms. *BioMed Res Int* **2015,** 1–14 (2015).

43.    Emiliani, S. *et al.* Mutations in the tat Gene Are Responsible for Human Immunodeficiency Virus Type 1 Postintegration Latency in the U1 Cell Line. *J Virol* **72,** 1666–1670 (1998).

44.    Sattentau, Q. J. & Stevenson, M. Macrophages and HIV-1: An Unhealthy Constellation. *Cell Host Microbe* **19,** 304–310 (2016).

45.    Gerngross, L. & Fischer, T. Evidence for cFMS signaling in HIV production by brain macrophages and microglia. *J Neurovirol* **21,** 249–256 (2014).

46.    Crowe, S. M. & Sonza, S. HIV-1 can be recovered from a variety of cells including peripheral blood monocytes of patients receiving highly active antiretroviral therapy: a further obstacle to eradication. *J Leukoc Biol* **68,** 345–350 (2000).

47.    Gavegnano, C. & Schinazi, R. F. Antiretroviral therapy in macrophages: implication for HIV eradication. *Antivir Chem Chemother* **20,** 63–78 (2009).

48.    Gavegnano, C. *et al.* Cellular Pharmacology and Potency of HIV-1 Nucleoside Analogs in Primary Human Macrophages. *Antimicrobl Agents Chemother* **57,** 1262–1269 (2013).

50.    Baxter, A. E. *et al.* Macrophage Infection via Selective Capture of HIV-1-Infected CD4. *Cell Host Microbe* **16,** 711–721 (2014).

51.    Rappaport, J. & Volsky, D. J. Role of the macrophage in HIV-associated neurocognitive disorders and other comorbidities in patients on effective antiretroviral treatment. *J Neurovirol* **21,** 235–241 (2015).

52.    Robillard, K. R. *et al.* Role of P-Glycoprotein in the Distribution of the HIV Protease Inhibitor Atazanavir in the Brain and Male Genital Tract. *Antimicrob Agents Chemother* **58,** 1713–1722 (2014).

53.    Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* **39,** D1005–D1010 (2010).

**Supplemental References**

1.    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 USA* **111,** E1843–E1851 (2014).

2.  Williams, M. V. & Parris, D. S. Characterization of a herpes simplex virus type 2 deoxyuridine triphosphate nucleotidohydrolase and mapping of a gene conferring type specificity for the enzyme. *Virology* **156,** 282–292 (1987).

3.  Seamon, K. J. *et al.* Small Molecule Inhibition of SAMHD1 dNTPase by Tetramer Destabilization. *J. Am. Chem. Soc.* **136,** 9822–9825 (2014).

4.  Weil, A. F. *et al.* Uracil DNA glycosylase initiates degradation of HIV-1 cDNA containing misincorporated dUTP and prevents viral integration. *Proc Natl Acad Sci USA* **110,** E448–E457 (2013).

5.  Seiple, L. A., Cardellina, J. H., Akee, R. & Stivers, J. T. Potent inhibition of human apurinic/apyrimidinic endonuclease 1 by arylstibonic acids. *Mol Pharmacol* **73,** 669–677 (2008).

6.  Kennedy, E. M. *et al.* Ribonucleoside triphosphates as substrate of human immunodeficiency virus type 1 reverse transcriptase in human macrophages. *J Biol Chem* **285,** 39380–39391 (2010).

7.  Fromentin, E., Gavegnano, C., Obikhod, A. & Schinazi, R. F. Simultaneous Quantification of Intracellular Natural and Antiretroviral Nucleosides and Nucleotides by Liquid Chromatography−Tandem Mass Spectrometry. *Anal Chem* **82,** 1982–1989 (2010).

8.  Jordan, A., Bisgrove, D. & Verdin, E. HIV reproducibly establishes a latent infection afteracute infection of T cells in vitro. *The EMBO Journal* **22,** 1868–1877 (2003).

9.  Sanders, R. *et al.* Evaluation of Digital PCR for Absolute DNA Quantification. *Anal Chem* **83,** 6474–6484 (2011).

10. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv* 1–9 (2012).

11. Ho, Y.-C. *et al.* Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell* **155,** 540–551 (2013).

12. Popescu, I. *et al.* Activation-induced Cell Death Drives Profound Lung CD4 +T-Cell Depletion in HIV-associated Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **190,** 744–755 (2014).

13. Sabbatucci, M. Endogenous CCL2 neutralization restricts HIV-1 replication in primary human macrophages by inhibiting viral DNA accumulation. *Retrovirology* **12,** 1–22 (2015).

14. Ellery, P. J. *et al.* The CD16+ Monocyte Subset Is More Permissive to Infection and Preferentially Harbors HIV-1 In Vivo. *J Immunol* **178,** 6581–6589 (2007).

15. Stopak, K., de Noronha, C., Yonemoto, W. & Greene, W. C. HIV-1 Vif Blocks the Antiviral Activity of APOBEC3G by Impairing Both Its Translation and Intracellular Stability. *Mol. Cell* **12,** 591–601 (2003).

16. Miyahara, S. *et al.* Discovery of a Novel Class of Potent Human Deoxyuridine Triphosphatase Inhibitors Remarkably Enhancing the Antitumor Activity of Thymidylate Synthase Inhibitors. *J Med Chem* **55,** 2970–2980 (2012).

# Erik Christian Hansen
30 East Preston Street, Apt 2
Baltimore, MD 21202
Phone: +01.206.755.5942
E-mail: erikchansen@gmail.com

## Education
**Ph.D., Chemical Biology; Johns Hopkins University, Baltimore, MD**
Anticipated graduation date: May 2016
Title: "Uracilated HIV-1 DNA follows diverse fates during infection of myeloid lineage cells"
Mentor: Dr. James T. Stivers

**M.S., Chemical Biology; Johns Hopkins University, Baltimore, MD**      **June 2013**

**B.A., Biochemistry; University of Colorado, Boulder, CO**      **August 2005**

## Summary of Qualifications
- Cell culture - primary cell isolation (healthy and HIV+ donors) under BSL-3 conditions, transient transfection/nucleofection, multi parameter flow cytometry, qPCR and ddPCR
- Virology - production of lentiviral vectors, virus, purification, quantification and viral outgrowth assays
- Protein expression - (*E. coli*, baculovirus) and purification (native/tagged), FPLC (ÄKTA, Bio-Rad platforms)
- Protein characterization - chromatography, SDS-PAGE, western blot, ELISA, silver stain, UV spectrometry and dynamic light scattering
- Enzyme kinetics - classical and allosteric enzymes (steady-state and single-turnover), fluorescence methods (anisotropy, FRET)
- Structural biology - protein X-ray crystallography
- Mass spectrometry - MALDI-ToF and LC-MS/MS
- Molecular biology - PCR, subcloning and sequencing

## Professional Experience
*Patient Coordinator* , **The Moore Clinic for HIV Care**     *August 2015 – present*
**Johns Hopkins Medicine, Baltimore, MD**
Serve as patient coordinator and petty cash custodian for clinical study entitled: 'Persistence and fate of invisible U/A base pairs in HIV-1 proviral DNA.' (IRB00038590).
- Contact research study participants, obtain informed consent and disburse funds.

***Research Associate III***                                    ***August  2007  –  August 2011***
***Assay Development and Screening***
**Catalyst Biosciences Inc., South San Francisco, CA**
Primary responsibilities included designing, optimizing and implementing *in vitro and* ex *vivo* biochemical assays to screen, profile and characterize Catalyst's engineered coagulation proteases.

- Experimental design, execution and evaluation of an ELISA based assay using a Biomek-Fx workstation to determine the pharmacokinetic parameters for protease variants in the discovery pipeline.
- Designed and implemented an alpha screen based assay for the detection and quantification of proteases displayed on phage. These efforts allow for kinetic parameters to be measured without the need for further-bacterial based amplification, significantly increasing throughput for Catalyst's therapeutic protease platform technologies.
- Established that partially purified supernatants from cultures could be used to quickly assess the kinetic properties for wild type and mutant proteases. The work determined that His/Fc-tagged proteases were kinetically indistinguishable from their untagged counterparts – allowing purification handles that have substantially accelerated the team's hit-to-protein cycle time.
- Relieved a key bottleneck in the discovery platform by developing methods for purifying and labeling biotinylated serine protease inhibitor (Serpin) baits and their detailed characterization.
- Charged with IDBS ActivityBase (a data capture, analysis and management software for maintenance of routine assay data) administration and development.

***Research Associate, Structural Biology Group***          ***July 2005 – August 2007***
**Beryllium (formerly deCODE Biostructures), Bainbridge Island, WA**
Supported the method development and execution of protein purification experiments at analytical and preparative scale for *in vitro* assays and X-ray crystallography. Utilized biochemical and biophysical approaches to study the structure and function of proteins and protein complexes.

- Executed the purification, crystallization and analysis for the structure determination of a viral protein fusion complex. Dynamic light scattering and mass spectrometry were used for the structural characterization and stoichiometric determination of RSV heterotrimers to achieve a stable hexameric coiled-coil complex. Co-crystal structures expanded client's understanding of the structure activity relationships for their inhibitors.
- Served as group leader for co-crystallization experiments in the design of a novel fragment based approach of using small molecule metabolites (substrates, co-factors, modulators, etc.) for identifying lead inhibitors that target previously unidentified allosteric binding sites in highly regulated proteins.

## Teaching and Mentoring Experience
### *Graduate Student Instructor, Organic Chemistry, JHU*     *August 2012 – June 2013*
○ Lead a 1.5-hour discussion for approximately 30 students each week.
○ Prepared summary lectures, problem sets and administered/graded exams, worked with students in groups and on an individual basis.

### *Graduate Student Instructor, Organic Mechanisms in Biology, JHU*    *2012 & 2013*
• Graded class assignments and exams.

### *Graduate Student Mentor, Chemical Biology Interface program, JHU 2012 – 2016*
• Mentored 5 graduate rotation students.

## Awards and Honors
• *Best poster*, Johns Hopkins Chemical Biology Interface (CBI) Program 11th Annual Retreat; September 2015.
• *St. Jude Future Fellow Research Conference*. Selected for all paid invitational conference. St Jude Children's Research Hospital; June 2015.
• *Scheinberg Travel Award*. Johns Hopkins Medical Institute; May 2015.
• *Best poster*, Johns Hopkins Pharmacology and Molecular Sciences Retreat; October 2014.
• *Student invited speaker*, Johns Hopkins CBI Program 9th Annual Retreat; October 2013.
• NIH Chemical Biology Training Grant, Johns Hopkins University; July 2011 – July 2013.
• *Member*, American Association for the Advancement of Science; August 2013.
• *SPoT (Superior Performance or Technology) award*, Experimental design and implementation of an ELISA based method for detecting engineered FIX/FIXa proteases to support preclinical mouse studies, Catalyst Biosciences Inc.; 2010.
• *SPoT award*, Developed a high-throughput biochemical assay for rapid characterization of mutant proteases for platform technology, Catalyst Biosciences Inc.; 2009.

## Publications
• **Hansen EC**, Ransom M, Hesselberth JR, Hosmane NN, Capoferri AJ, Zhang H, Bruner KM, Pollack RA, Drummond MB, Siliciano JM, Siliciano RF, Stivers JT. Diverse fates of uracilated HIV-1 DNA during infection of myeloid lineage cells. *Nature Comm*, submitted.

• Seamon, KJ, **Hansen EC**, Kadina AP, Kashemirov BA, McKenna CE, Bumpus NN and Stivers JT. (2014) Small molecule inhibition of SAMHD1 dNTPase by tetramer destabilization. *J Am Chem Soc*. (2014) 136(28): 9822-9825.

• **Hansen, EC**, Seamon, KJ, Cravens, SL and Stivers, JT (2014). GTP activator and dNTP substrates of HIV-1 restriction factor SAMHD1 generate a long-lived activated state. *Proc Natl Acad Sci USA* (2014) 111(18): E1843–E1851.

- Davies, D, Stewart, L., Mamat, B., Magnusson, O., Christensen, J., Haraldsson, M., Pease B., **Hansen, EC**, Singh J, Zembower D, Kim H, Kiselyov AS, Burgin AB, Gurney ME, Stewart L. Discovery of leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography. *J Med Chem* (2009) 52, 4694-4715.

## Seminars and Poster Presentations

- **Hansen EC**, Siliciano RF and Stivers JT. Invisible U/A base pairs within HIV DNA: Diverse fates during infection of monocyte-derived macrophages. *Poster*. Johns Hopkins CBI Program 11th Annual Retreat. September 2015. Baltimore, MD.

- **Hansen EC**, Ransom M, Hesselberth JR, Hosmane NN, Siliciano JM, Silciano RF and Stivers JT. dUTP and Uracil Base Excision Repair (UBER) in HIV-1 infection of macrophages. *Poster*. St. Jude's Future Fellow Research Conference. June 2015. Memphis, TN.

- **Hansen EC**, Ransom M, Hesselberth JR, Hosmane NN, Siliciano JM, Silciano RF and Stivers JT. dUTP and Uracil Base Excision Repair (UBER) in HIV-1 infection of macrophages. *Seminar*. Cold Spring Harbor Laboratory Conference, Retroviruses. May 2015. Long Island, NY.

- **Hansen EC**, Siliciano RF and Stivers JT. Mapping the Spatial Distribution and Temporal Fate of 'Invisible' U/A base pairs in unintegrated and integrated HIV-1 DNA. *Poster*. Johns Hopkins CBI Program 10th Annual Retreat. October 2014. Baltimore, MD.

- **Hansen EC**, Siliciano RF and Stivers JT. Mapping the Spatial Distribution and Temporal Fate of 'Invisible' U/A base pairs in unintegrated and integrated HIV-1 DNA. *Poster*. Johns Hopkins Pharmacology and Molecular Sciences Annual Retreat. October 2014. Baltimore, MD.
- Boucher LE, **Hansen EC**, Stivers JT and Bosch J. Identifying Novel Invasion Inhibitors of P. Falciparum. *Poster*. Johns Hopkins CBI Program 9th Annual Retreat. October 2013. Baltimore, MD.