Published in final edited form as:

J Am Chem Soc. 2019 May 22; 141(20): 8327–8338. doi:10.1021/jacs.9b02743.

Inhibition of HIV Maturation via Selective Unfolding and Cross-Linking of Gag Polyprotein by a Mercaptobenzamide Acetylator

Lisa M. Miller Jenkins[†], Elliott L. Paine[†], Lalit Deshmukh^{‡,§}, Herman Nikolayevskiy[§], Gaelyn C. Lyons[†], Michael T. Scerba[§], Kara George Rosenker[§], Hans F. Luecke^{||}, John M. Louis[†], Elena Chertova[⊥], Robert J. Gorelick[⊥], David E. Ott[⊥], G. Marius Clore[‡], Daniel H. Appella^{*,§}

[†] Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States

[‡] Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, United States

[§] Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, United States

^{||} Advanced Mass Spectrometry Core, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892, United States

[⊥] AIDS and Cancer Virus Program, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21701, United States

Abstract

For HIV to become infectious, any new virion produced from an infected cell must undergo a maturation process that involves the assembly of viral polyproteins Gag and Gag–Pol at the membrane surface. The self-assembly of these viral proteins drives formation of a new viral particle as well as the activation of HIV protease, which is needed to cleave the polyproteins so that the final core structure of the virus will properly form. Molecules that interfere with HIV maturation will prevent any new virions from infecting additional cells. In this manuscript, we characterize the unique mechanism by which a mercaptobenzamide thioester small molecule (SAMT-247) interferes with HIV maturation via a series of selective acetylations at highly conserved cysteine and lysine residues in Gag and Gag–Pol polyproteins. The results provide the first insights into how acetylation can be utilized to perturb the process of HIV maturation and reveal a new strategy to limit the infectivity of HIV.

* Corresponding Author : appellad@niddk.nih.gov .

Present Address : Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093 (L.D.).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/jacs.9b02743](https://doi.org/10.1021/jacs.9b02743).

Synthetic procedures, characterization of synthetic compounds, sites of reaction identified by mass spectrometry, sequence conservation of modified lysines, qRT-PCR data on RNA packaging, analysis of effects on protease activity, data on thioester–thiol exchange (PDF)

Notes

The authors declare no competing financial interest.

Graphical Abstract

INTRODUCTION

Antiretroviral drugs targeting HIV enzymes (namely, protease, reverse transcriptase, and integrase) effectively reduce detectable viremia and promote long-term survival in patients. However, even in patients with no detectable viremia upon treatment with state-of-the-art combined antiretroviral therapy (cART), the virus rapidly recrudesces after withdrawal of therapy. Therefore, strict adherence to, as well as lifelong treatment with, antiretroviral medication is needed for HIV-infected patients to prevent the development of AIDS symptoms. Long-term cART treatment results in drug-induced toxic effects as well as emergence of HIV drug resistance due to failure of patients to adhere to treatment regimens. Disturbingly, increased HIV resistance to frontline treatments has been detected in newly diagnosed infections.^{1,2} Therefore, new HIV inhibitor targets and mechanisms are needed, especially those that pose a significant obstacle for the fitness of viral resistance mutants.

Assembly of the virion in infected cells is a critical step in HIV replication and is driven by the Gag polyprotein, consisting of the matrix (MA), capsid (CA), nucleocapsid (NC), and p6 proteins linked together via peptide spacer sequences. Gag polyproteins selectively package the HIV genomic RNA dimer through two strictly conserved Cys–Cys–His–Cys zinc knuckles of the NC protein. Produced by a frameshift at the NC/p6 junction that occurs in ~5% of Gag translations, the Gag–Pol polyprotein encodes all retroviral enzymes: integrase (IN), protease (PR), and reverse transcriptase (RT). Viral assembly begins at the plasma membrane where the viral structural proteins, primarily the Gag and Gag–Pol polyproteins, condense to form nascent immature particles that bud through the plasma membrane and release from the cell.³ During or just after budding, the Gag and Gag–Pol polyproteins in the immature particle are cleaved by the viral protease resident in Gag–Pol, which is self-activated via transient dimerization with an adjacent Gag–Pol protease domain. The intermediate steps of HIV-1 protease autoprocessing rely on transient, sparsely-populated encounter complexes to accomplish the initial intramolecular cleavage at the junction between p6 and protease in Gag–Pol.⁴ Subsequent intermolecular cleavage by HIV protease of both Gag and Gag–Pol induces conformational changes that culminate in the assembly and formation of the mature virion structure, which is required for particle infectivity. This maturation process is very sensitive to perturbations in both the timing and ordered cleavage of the polyproteins and is a prime target for antiviral intervention.^{5,6}

The mercaptobenzamide thioester SAMT-247 (**1**, Figure 1A) reacts with the strictly conserved Cys–Cys–His–Cys knuckles of NC. Extensive site-directed mutagenesis studies of NC have shown that these exact residues and their spacing are strictly required for orthoretrovirus infectivity.^{7,8} Compound **1** exploits the weaker coordination of Zn²⁺ by the C-terminal knuckle.⁹ The result of weaker Zn²⁺ coordination is that at any one moment, approximately 2% of the C-terminal knuckle has one of its cysteine side chains in an uncoordinated sulfhydryl form, allowing reaction with **1**.¹⁰ Acetylation of this cysteine at the C-terminus of HIV NC ejects the zinc ion, leading to its irreversible unfolding.^{11–13} Either before or concomitant with knuckle unfolding, the acetyl group is transferred to an adjacent lysine.¹³ Thus, the net action of **1** is to acetylate both cysteines and lysines in NC. This mechanism of inhibition by covalent acetylation has also been demonstrated for targeting the androgen receptor, a therapeutic target for advanced prostate cancer.¹⁴

After acetylating NC in cells, **1** is converted to the thiolate **2**, which can be reacylated by acetyl coenzyme A to regenerate the active thioester **1**, which, in turn, can attack another molecule of NC (Figure 1A).¹³ This regenerative cycle is likely to result in acetyl transfers to proteins other than NC, yet the very low toxicity of **1** and similar mercaptobenzamides suggests that this catalytic cycle of acetylation is not overtly detrimental to the cell.¹⁵ Although **1** can acetylate the cysteines and lysines of HIV NC in vitro and in infected cells, we hypothesized that **1** would likely acetylate other residues of the Gag polypeptide and that any such additional acetylation sites would reveal how **1** is able to exert antiviral activity without evidence of viral resistance to this molecule. To examine the extent of Gag and Gag–Pol protein acetylation, we combined mass spectrometric techniques with cell biology and biophysical approaches to study the effect of **1** on Gag. The results reveal that the antiviral mechanism of **1** is significantly more complex than the previously reported reactions with NC. In particular, **1** is able to acetylate multiple, highly conserved cysteine and lysine side chains on Gag following the basic reactivity patterns studied in a kinetic analysis of **1** reacting with model compounds. The acetylation pattern of Gag promoted by **1** is consistent across both in vitro- and cell-based studies, demonstrating that the reactivity of **1** should be used as a new approach to inhibit HIV maturation via this unique acetylation-based mechanism.

RESULTS

Modifications of NC Lysines by **1** and Quantification by Mass Spectrometry.

Mass spectrometry analysis was performed to map the sites of NC modified by **1** in solution. After incubation of NC with a 10-fold molar excess of **1** at 30 °C for 16 h, the protein was digested with either endoproteinase Glu-C or Arg-C to obtain full sequence coverage of the lysine-rich protein. Mass spectrometry identified each lysine residue in NC that showed evidence of modification by **1** (Figure 1B and Supplemental Table 1) with distinct differences in the amount of acetylation at the specific locations. The most heavily acetylated sites are at lysines in the C-terminal knuckle (NC numbering: Lys38, Lys41, Lys47; Gag numbering: Lys415, Lys418, Lys424), as compared to the N-terminal knuckle (NC Lys20; Gag Lys397), consistent with previous NMR analyses demonstrating that the C-terminal knuckle is the site of initial reaction.^{10,13} Interestingly, the lysines in the flexible

peptide region between the two zinc knuckles (NC Lys33 and Lys34; Gag Lys410 and Lys411) and just prior to the N-terminal knuckle (NC Lys11 and Lys14; Gag Lys388 and Lys391) are acetylated the least by comparison. Collectively, these results indicate that the C-terminal knuckle is more prone to acetylation by **1** compared to the N-terminal knuckle, consistent with the initial attack of the C-terminal knuckle and subsequent induction of NC unfolding contributing to the strong preferences for specific sites. As we have previously shown that the Zn^{2+} coordination of NC becomes unstable during the reaction with **1**,^{10,11} it is likely that some of the observed acetylation occurs subsequent to the loss of Zn^{2+} coordination when the protein is in a different conformational state.

Reaction of **1** with HIV Gag Monitored by NMR.

Several studies demonstrated that **1** is able to acetylate NC in vitro and promote the loss of Zn^{2+} coordination,^{10–13} but it was unclear whether the same reaction occurs between **1** and NC within HIV Gag when the domains of matrix and capsid are covalently attached in a complete polyprotein of viral assembly. To analyze the effects of **1** on Gag, we monitored the protein conformation by NMR spectroscopy by adding excess **1** to recombinant Gag polyprotein containing MA, CA, SP1, and NC domains (Gag p6). The addition of **1** to Gag p6 initially shows decreases in $^1\text{H}/^{15}\text{N}$ cross-peak intensities at the C-terminal knuckle of the NC domain (Figure 2). Additional decreases in cross-peak intensities are subsequently observed in the N-terminal knuckle of NC, followed by loss of NMR signal intensity in MA and CA. It should be noted that the cumulative peak heights and peak volumes as a function of time are the same, indicating that the reduction in peak height is due to disappearance of signal as a result of aggregation, as opposed to broadening arising from an exchange phenomenon, in which case the peak heights would decrease, but the volumes would remain unchanged. Thus, these findings indicate that **1** reacts with Gag p6, likely promoting aggregation of the entire Gag polyprotein.

To confirm that **1** reacts with Gag p6 via an acyl transfer mechanism, we synthesized a pentynoyl derivative (**S1**; Supplemental Figure 1) to probe for transfer of the pentynoyl group to the protein. Incubation of **S1** with Gag p6 followed by a copper-catalyzed click reaction with rhodamine azide clearly demonstrated efficient labeling of the Gag p6 protein (Supplemental Figure 1). Thus, it is clear that acyl mercaptobenzamides are able to react with Gag p6 polyprotein in solution via transfer of the acyl group on the molecule to the protein.

Mass Spectrometry Mapping of Acetylation Sites in Gag in Vitro after Reaction with **1**.

Mass spectrometry analysis was performed to identify the sites of covalent reaction with Gag p6. Following incubation of Gag p6 with a 10-fold molar excess of **1** at 30 °C for 16 h, the protein was digested with either trypsin or endoproteinase Arg-C, with the endoproteinase Arg-C used to enable sequence coverage of the lysine-rich NC domain. As shown in Supplemental Figure 2A, the analysis yielded nearly complete sequence coverage of Gag p6. Similar to NC, all of the cysteine and lysine residues within the NC region of Gag p6 were found to be modified by **1** (Supplemental Table 2). A specific pattern of cysteine and lysine acetylation was observed throughout the Gag protein, with extensive acetylation observed in MA and NC domains and some acetylation in the C-terminal

domain of CA (Supplemental Table 2 and Supplemental Figure 2A). Very small amounts of acetylation were present in the N-terminal domain of CA. Label-free quantitation of the cysteine-modified peptides that were commonly observed after 4 and 24 h of reaction with **1** demonstrated an increase in cysteine modification over time (Table 1). Surprisingly, **1** did not result in the complete acetylation of every cysteine or lysine in the protein, suggesting a structure-driven preference for the reaction between **1** and Gag p6.

To better understand the importance of the sites of reaction, the stoichiometry of modification was determined. Gag p6 was incubated with a 10-fold molar excess of **1** at 30 °C for either 4 or 24 h. After the reaction with **1**, the protein was treated with ¹³C-labeled acetic anhydride (“heavy”), which labeled all the remaining unmodified lysine residues with a ¹³C-labeled acetyl group. The relative proportions of “light” acetylation, corresponding to reaction with **1**, and “heavy” acetylation, corresponding to no reaction with **1**, were determined from the intensities of the immonium ions observed in the MS/MS fragment spectrum.¹⁶ Following 4 h of incubation with **1**, three primary regions of Gag p6 modification were observed—at Lys30 and Lys32 in MA, at Lys157 and Lys162 in CA, and Lys388 and K391 in NC (Table 2). These modifications increased substantially after 24 h, with new modification sites within MA, CA, and NC observed. At 24 h, the greatest relative proportion of reacted protein was observed for lysine residues at the N-terminus of MA and at Lys290 in CA.

Probing the Reaction Selectivity of **1** with Cysteine, Followed by Intramolecular Transfer.

Our MS data suggested that **1** did not acetylate Gag cysteines and lysines indiscriminately; rather, it demonstrated spatial and temporal preferences for certain residues within the Gag polyprotein. Therefore, we hypothesized that the observed acetylation pattern was a consequence of sequential *intramolecular* acetyl transfers from reactive cysteine thioesters (acetylated by **1**) to neighboring lysine residues. To experimentally model the effect of cysteine–lysine distance on the rate of intramolecular acetyl transfer, four aminothiols of increasing N–S distance [cysteine (**3a**), homocysteine (**4a**), the Cys–Gly dipeptide **5a**, and the Cys–β-Ala dipeptide **6a**] were treated with **1** under identical experimental conditions (solvent = 93:7 PBS 10×/D₂O, pH = 7.4 ± 0.3, temperature = 25 °C, [SAMT-247(**1**)] = [aminothiol] = 1.5 mM; Figure 3A) and monitored by ¹H NMR spectroscopy over the course of ~12 h. All four reactions followed a similar course—initial *intermolecular* acetylation at sulfur, generating the intermediate thioesters **3b**, **4b**, **5b**, and **6b**, was followed by an *intramolecular* acetyl transfer to nitrogen, producing the acetamides **3c**, **4c**, **5c**, and **6c** (Figure 3A). Although the measured rate of *intermolecular* thiol acetylation proved comparable for all aminothiols (**3a**–**6a**) [Figure 3B; k_1 (**3a** + **1**): 24 ± 4 M⁻¹ s⁻¹, k_2 (**4a** + **1**): 23 ± 3 M⁻¹ s⁻¹, k_3 (**5a** + **1**): 85 ± 5 M⁻¹ s⁻¹, k_4 (**6a** + **1**): 88 ± 8 M⁻¹ s⁻¹], the rate of *intramolecular* thioester–amine exchange varied greatly, resulting in different observable abundances of the intermediate thioesters **3b**, **4b**, **5b**, and **6b** (Figure 3C). Aminothiols with larger S–N distances (**5a** and **6a**) achieved high, persistent concentrations of the intermediate thioesters **5b** (maximum conversion of 87%) and **6b** (full conversion), indicating slow *intramolecular* acetyl transfer from sulfur to nitrogen. Conversely, for aminothiols with smaller S–N distances (**3a** and **4a**), the observable abundance of the intermediate thioesters was very low (**3b**: completely unobservable; **4b**: < 5%), indicating rapid *intramolecular*

acetyl transfer to generate the acetamides **3c** and **4c**. Our data suggests that the rate of acetamide (**3c**, **4c**, **5c**, and **6c**) formation strongly depends on the average thioester–amine distance (**6b** > **5b** > **4b** > **3b**), which determines the entropic barrier to formation of the acetyl transfer transition state. When applied in the context of the Gag polypeptide, this model predicts that cysteines acetylated by **1** would trans-acetylate more quickly to proximal lysines, whereas lysines without spatially proximal cysteines would remain unacetylated. Therefore, the selectivity of Gag acetylation by **1** is predicted by the differences in model compound reactivity.

Mass Spectrometry Mapping of Acetylation Sites in Gag in Virions after Treatment of HIV-Infected Cells with **1**.

To confirm that the acetylated lysines identified in vitro are also acetylated in vivo, we next analyzed the sites of modification on Gag isolated from virions released by infected CD4 T cells treated with **1**. For these studies, the chronically HIV_{MN}-infected H9 CD4 T cell line, Clone 4,¹⁷ was used, because it constitutively produces high levels of infectious HIV_{MN} virions with reproducible properties.¹⁸ To detect modified residues, the proteins from the viral particles had to be purified by liquid chromatography prior to trypsin digestion. Because of the potential for changes in the cysteine modifications during the purification, the cysteine residues were reduced and alkylated before mass spectrometry analysis. Therefore, only lysine modifications can be analyzed in this manner.

Similar to the sites of reaction identified by mass spectrometry in the in vitro reaction, the Gag in virions produced from **1**-treated cells contained multiple acetylated lysine residues throughout the MA, CA, NC, and p6 domains (Table 3, Supplemental Figure 2B, and Supplemental Table 3). In addition, five sites, Lys76, Lys103, and Lys113 in MA, Lys293 in CA, and Lys488 in p6 (all numbering for HIV_{MN}) were observed to be acetylated in the untreated cells, indicating that these are basal modifications of Gag. To the best of our knowledge, this represents the first observation of cellular acetylation of HIV Gag. As shown in Table 3, lysine residues at the beginning and end of MA were identified as sites of reaction both in vitro and from infected cells. More sites of reaction in CA were identified from treated cells, and multiple sites of modification within the p6 domain of Gag, which was not present in the in vitro construct, were also identified. Thus, the results from virions released by treated cells largely recapitulate the sites of reaction observed in vitro. Of the sites that were modified by reaction with **1**, 12 are conserved as lysine residues among at least 85% of HIV strains (Supplemental Figure 3). This high degree of conservation suggests the importance of not only a charged residue at these positions but particularly for a lysine. As modification by **1** would neutralize the charge and lengthen the side chain at these positions, it is possible that these modifications would have some effect on replicative fitness. As can be seen in a model of Gag p6 generated using the coordinates of the isolated MA, CA, and NC domains,¹⁹ these sites are solvent exposed and available to make interactions (Figure 4).

Functional Consequence of Gag Acetylation by **1**.

To examine the structures of virions treated with **1**, we used transmission electron microscopy to visualize HIV-1 virions produced from chronically infected cells in the

presence or absence of 100 μ M **1**. The images show that in contrast to the untreated virions with their classic cone shaped morphology, the virions released after treatment with **1** had severe assembly defects that fell into two general categories (Figure 5A). In the first and most prevalent type (labeled 1 in Figure 5A), the virions have the morphology of a protease mutant, which contains a layer of unprocessed Gag apparent as a discontinuous electron dense bar that covers most of the interior membrane of the virion, as we have previously observed.²⁰ The second and less frequent type possesses a poorly condensed circular core (Figure 5B), similar to SP1 (or Spacer 1) mutants.²¹ Taken together, these data are consistent with both complete and partial blocks to protease processing of Gag and Gag–Pol.

As the NC domain of Gag is essential for packaging of viral genomic RNA into virions, we next examined whether treatment with **1** affected genomic RNA levels in virions. The amount of genomic RNA was quantified from HIV virions produced from HEK293T cells transfected with the pNL4–3 infectious molecular clone in the presence or absence of 100 μ M **1** sampled 32–72 h post-transfection/**1** treatment. Virion-associated genomic RNA was measured by a qRT-PCR assay.²² The assay results, shown in Supplemental Figure 4, detected robust levels of genomic RNA in all of the samples, with little difference between the levels of genomic RNA in the untreated and **1**-treated samples. Thus, treatment of cells with **1** does not lead to a defect in viral RNA packaging. Because the zinc knuckles are strictly required for genomic packaging, these results imply that **1** attacks and ejects zinc later in the assembly process, sometime after genome-specific recognition.

To better understand the molecular consequences of treatment with **1**, we analyzed the proteins in the released viral particles by Western blotting. Previously, we showed that treatment of infected cells with **1** resulted in accumulation of uncleaved Gag, which was cross-linked by disulfide bonds.¹³ Separating the viral particles under reducing conditions and blotting with antibodies against CA and NC confirmed that treatment with **1** results in a processing defect, as significant amounts of unprocessed Gag were detected (Figure 5C). Separation under nonreducing conditions also showed that the Gag exists as high-molecular-weight aggregates (Figure 5D), consistent with Gag multimerization, likely arising from intermolecular cross-links between cysteine residues in the NC domain of Gag following zinc ejection. Western blots using antibodies against RT demonstrate that, like Gag, Gag–Pol is not processed fully in cells treated with **1** (Figure 5C). Blotting for RT using nonreducing conditions showed that RT and Gag–Pol also exist as high-molecular-weight aggregates (Figure 5D). As there is approximately 20 times more Gag than Gag–Pol in infected cells,²³ it is likely that much of the aggregated Gag–Pol is cross-linked to Gag molecules.

1 Does Not Affect HIV Protease Activity.

As the primary effect of **1** on virions was a defect in polyprotein Gag processing that could arise due to inhibition of HIV protease, we examined whether **1** inhibited protease (PR) per se by analyzing autoprocessing and mature PR activities using in vitro assays. First, we used a model precursor for Gag–Pol composed of the transframe region (TFR) fused to PR (TFR–PR) to assess the effect of **1** on N-terminal autoprocessing mediated by PR (Supplemental Figure 5A). Whereas cleavage of the TFR at the N-terminus of PR was complete at nearly 10 min in the absence of an inhibitor, treatment with the specific PR inhibitor darunavir

(DRV) completely inhibited the reaction even after 30 min (Supplemental Figure 5B,C). In contrast, treatment with **1** had no effect up to 50 μ M for 30 min on transient dimerization, which is crucial for N-terminal autoprocessing of the protease precursor (TFR-PR) and release of the mature protease (Supplemental Figure 5B,C). This cleavage (TFR/PR) is concomitant with stable dimer formation and appearance of mature protease-like catalytic activity. No change in either the disappearance of the precursor or the appearance of the products is observed in the presence of **1** relative to the control reaction.

Finally, we used mass spectrometry to identify any potential reaction sites of **1** with PR isolated from virions released by treated, infected cells. Although we had reasonable sequence coverage of PR (Supplemental Figure 6), we did not observe any modified residues. We did, though, identify one site in RT (Lys719) and one in IN (Lys1341) that were modified (Supplemental Table 3). It should be noted that these identifications likely do not originate from Gag-Pol but from the cleaved proteins, as these identifications originated from proteins with molecular weights consistent with cleaved protein and not intact Gag-Pol. Overall, our results suggest that the processing defects observed in cells treated with **1** are not the result of PR inhibition.

DISCUSSION

The ability of HIV-1 to develop resistance against inhibitors is a defining characteristic of the virus that complicates the development and implementation of therapies to control the infection. Potential antiretroviral drugs that target viral assembly, other than the enzymatic protease inhibitors, are few. Candidate molecules such as betulinic acid derivatives (bevirimat) and the recently reported GS-CA1 interfere with ordered protease cleavage of Gag via binding to structural elements of Gag; however, HIV develops resistance to these molecules in a manner analogous to protease inhibitors.^{24,25} Exploring alternative strategies to inhibit HIV, for example, through the acetylation-promoted cross-linking of Gag, may eliminate the ability of HIV to develop viral resistance and lead to effective treatments with simpler regimens. Understanding the mechanistic details of how to effectively acetylate and interfere with the normal functions of HIV Gag while not interfering with regular cellular functions, as demonstrated by low toxicity in animal models, is crucial to advancing this novel strategy toward preclinical development.

The molecule **1** has a unique effect on HIV Gag that results from its ability to rapidly transfer an acetyl group to the thiol of cysteine side chains. The acetyl transfer reactivity of **1** takes advantage of a weak point in the HIV nucleocapsid protein: a transient, sparsely populated state of the protein where Cys36 (Gag Cys413) of NC becomes unligated to the zinc in the C-terminal knuckle.¹⁰ Acetylation of NC Cys36 followed by intramolecular acetyl transfer to adjacent lysine side chains destabilizes zinc coordination and disrupts the native structure of the protein.¹³ On the basis of our previous studies^{11,12} and the results obtained here, this is likely the key initial site of Gag disruption that is promoted by **1**. Once the C-terminal finger loses coordination to zinc, the N-terminal finger is also destabilized, followed by loss of the native structure of NC and exposure of several free cysteines. Additional acetylation at other sites in Gag facilitated by **1** may further interfere with proper maturation of the virus. The results in this manuscript clearly show that **1** modifies

multiple cysteines in Gag as well as several lysines, and these results were demonstrated both in vitro (for cysteine and lysine residues) and in cells (for lysine). Indeed, in vitro modeling of various molecular spacings between cysteine and lysine supports a proximity dependence for acetyl transfer and provides a partial mechanistic explanation for the lysine modification patterns observed. We hypothesize that the acetylated NC domain becomes disordered once zinc coordination is lost and promotes transacetylation of lysine residues from other domains of Gag, though acetylation at Gag Cys57 in matrix and Gag Cys330 and 350 in capsid, as identified in vitro, may be responsible for acetylation of lysines in these specific domains. Loss of zinc coordination in NC results in the exposure of free cysteine side chains that subsequently lead to disulfide cross-linking and the formation of Gag oligomers. The nature of disulfide-cross-linked Gag oligomers is likely heterogeneous, with disulfide bonds between many different Gag proteins cross-linked through the NC domains. Additional cross-linking may also involve disulfides between matrix and capsid domains.

Characterization of the antiviral activity of **1** by electron microscopy and Western blot analysis shows that the molecule interferes with proper Gag processing by protease, and the electron microscopy images of immature virions look similar to ones observed with protease deficient viruses.²⁰ Our results, though, clearly demonstrate that **1** is not a direct inhibitor of HIV protease itself. For HIV to properly mature, the protease domain of Gag–Pol must transiently dimerize so that HIV protease becomes active and proteolyzes Gag into MA, CA, NC, and p6 proteins. We hypothesize that cysteine residues exposed by reaction with **1** cause disulfide cross-linking between molecules of Gag and Gag–Pol; as there is estimated to be 20-fold more Gag than Gag–Pol,²³ most of the Gag–Pol would be disulfide-cross-linked with Gag, preventing formation of the Gag–Pol dimers required for protease autocatalytic activation (Figure 5E). If Gag–Pol is sequestered by disulfide cross-links with Gag promoted by **1**, then dimerization of protease domains in Gag–Pol can be prevented, and HIV protease activation will not occur.

Considering the nonspecific reactivity of **1** with cysteine thiols, it is remarkable that **1** can selectively interfere with HIV Gag processing and maturation without simultaneously interfering with other critical cellular functions. In this study and many previous experiments, cellular toxicity of **1** (as well as animal toxicity) is not observed at any concentration tested ($TC_{50} > 100 \mu M$ is typical for most cells, and in rats, the maximum tolerated dose is above 2 g per kg for a prodrug analogue of **1** when administered orally).^{15,26–29} The reactivity of **1** indicates that there would be multiple acetylations of cysteines in non-HIV proteins, and indeed, we previously showed that **1** can react with other zinc-finger proteins in vitro.³⁰ Consistent with this broad reactivity, the results in this study demonstrate that several cysteines in the MA and CA domains of Gag are acetylated by **1** along with a number of lysines in these domains. Although it is likely that **1** reacts similarly with other cellular proteins, the lack of toxicity in cells or animals treated with **1** and related molecules^{15,27–29} indicates that any erroneous acetylations of cellular proteins either occur on a minor population of a given protein or are easily corrected within cells by existing enzymes, such as deacetylases. It is also likely that the intracellular levels of **1** are very low under homeostatic conditions. Thus, we propose that the levels of **1** that are

present in a cellular environment will very quickly transfer an acetyl group to an available thiol to produce the thiolate **2**. Although the reaction of **2** with acetyl CoA will regenerate **1**, the thermodynamic equilibrium of this reaction highly favors the reactants (which are **2** and acetyl CoA). Therefore, **2** is likely the major species present in cells.

We further hypothesize that the ability of **1** to have an effect on HIV Gag in cells may be constrained to environments where there are very high local concentrations of Gag during the assembly and budding processes. In the immature virion, local concentrations of Gag approach 4 mM.³¹ In this situation, the presence of **1** as a minor species could have an enhanced ability to react with the dark state of NC and other sites in Gag. An initial reaction between **1** and Gag could set up a nucleation event causing the unfolding and disulfide cross-linking of multiple molecules of Gag and Gag–Pol present at high local concentrations at the cell surface because of budding of a new viral particle. Thus, the local high concentration of Gag and Gag–Pol promotes productive reaction with **1**. Furthermore, as the thiol **2** released by this reaction can be activated to reform the active thioester, the potential for a high level of reaction is increased. The final antiviral effect of these events is to interfere with the ability of HIV to properly mature.

Multiple attempts to generate HIV resistance to **1** have not been successful.^{15,32} The amino acid residues within NC, especially the cysteines coordinated to zinc, are strictly conserved across all HIV viruses, and the Cys–Cys–His–Cys array is found in all orthoretroviral NC proteins. Thus, unlike complex targets for antiretroviral agents such as enzymes and ligand binding sites, which have structural degrees of freedom with which to mutate around, the zinc-coordinated Cys–Cys–His–Cys array of NC is strictly required for function, leaving less flexibility for modification.⁷ Additionally, other cysteines in Gag are also highly conserved, as are a number of the lysines that are acetylated by **1** in both NC and CA. The ability of **1** to acetylate many highly conserved residues in Gag may be one reason that there is a high barrier to HIV developing resistance against this inhibitor in model systems. At the same time, the ability of **1** to acetylate a wide variety of cysteines may be a line of attack to which the HIV virus is not able to overcome with simple mutations.

The novel mechanism and reactivity of **1** presents new opportunities and strategies to inhibit the HIV virus. Developing molecules that covalently modify highly conserved residues in HIV Gag may lead to new therapeutics that avoid the issues associated with viral resistance.

EXPERIMENTAL SECTION

Materials.

Commercial solvents were obtained from Millipore Sigma. Commercial reagents were obtained from Millipore Sigma with the following exceptions: Fmoc–Cys(Trt)–OH (**S7**; Combi-Blocks); Boc–Gly–OH (Combi-Blocks); Boc– β -Ala–OH (Combi-Blocks); HOAt (Advanced ChemTech). Anhydrous commercial solvents were dispensed from a JC Meyer solvent drying system. Commercial reagents were used as received. Mercaptobenzamides **1** and **2** were both synthesized as described previously.³³

Protein Expression and Purification.

HIV-1 Gag p6 (MA-CA-SP1-NC, residues 1–432, plasmid HXB2) was subcloned in a pET-11a vector and expressed in BL21-CodonPlus (DE3)-RIPL competent cells (Agilent Technologies). Protein was expressed at 19 °C as described previously.¹⁹ Briefly, cells were grown at 37 °C in 1 L of Luria–Bertani (LB) medium at natural isotopic abundance or minimal M9 medium containing 0.3 g/L ²H/¹⁵N Isogro (Sigma-Aldrich), 99.9% (v/v) D₂O, 1 g/L ¹⁵NH₄Cl, and 3 g/L ²H₇, ¹²C₆-D-glucose for ²H/¹⁵N labeling. The temperature was reduced to 19 °C, 30 min prior to incubation. Cells, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at an optical density of A₆₀₀ ≈ 0.8, were harvested 24 h after induction. Protein was purified using a combination of ion exchange and size exclusion chromatography using our published protocol.^{19,34}

Cloning, expression, and isolation of the mature protease (PR) and its precursor (TFR-PR) have been described before.^{35,36} Induction for protein expression, isolation of inclusion bodies, and protein purification followed previously described protocols.^{37,38}

Thiol–Thioester Exchange Rate Constant Determination.

A solution of *tert*-butanol (233 nL, 2.45 μmol, 0.33 equiv) in PBS 10× buffer/D₂O (93:7, 4.9 mL) was added to either cysteine hydrochloride monohydrate (**3a**; 1.3 mg, 7.50 μmol, 1.00 equiv), homocysteine (**4a**; 1.0 mg, 7.50 μmol, 1.00 equiv), the Cys–Gly dipeptide **5a** (2.8 mg, 7.50 μmol, 1.00 equiv), or the Cys–β-Ala dipeptide **6a** (2.9 mg, 7.50 μmol, 1.00 equiv). Monitoring with a calibrated Orion Star A111 pH benchtop meter, the pH of the resulting mixture was adjusted to 7.40 ± 0.03 with aqueous sodium hydroxide solution (0.5 M in H₂O). An aliquot (~500 μL) of the pH-adjusted solution was transferred to an NMR tube, and the resulting sample was used to lock (onto D₂O) and shim the spectrometer. The NMR sample (500 μL) and a solution of SAMT-247 (**1**; 2.0 mg, 7.50 μmol, 1 equiv) in DMSO-*d*₆ (100 μL) were sequentially added to the reaction mixture, at which point a timer was started. Following an initial delay of ~60 s, a water-suppressed ¹H NMR spectrum at 25 °C was obtained at multiple time points. The acquisition of two dummy scans preceded the acquisition of each FID (eight scans), requiring 17 s per FID. For cysteine (**3a**) and homocysteine (**4a**), the time delays in between scans were as follows: 15 × 0.1 s, 2 × 30 s, 2 × 60 s, 2 × 120 s, 2 × 240 s, and 2 × 480 s. For the Cys–Gly dipeptide **5a** and the Cys–β-Ala dipeptide **6a**, the time delays in between scans were as follows: 15 × 0.1 s, 2 × 30 s, 2 × 60 s, 2 × 120 s, 2 × 240 s, 2 × 480 s, 4 × 900 s, 11 × 1800 s. After being referenced to the residual protium of DMSO-*d*₆, integration of the acetyl methyl group, δ 2.27 (s, 3H), relative to the integration of the *tert*-butanol internal standard, δ 1.04 (s, 9H), was used to calculate reactant concentrations. Because [**1**] ≈ [thiol] at early time points, data (the first 15 time points) was fit to the integrated form of the second order rate equation, $\frac{1}{[1]} - \frac{1}{[1]_0} = k\theta t$, derived from the second order rate equation, $-\frac{d[1]}{dt} = k[1][\text{thiolate}] = k\theta[1]^2$, where [**1**]₀ is the concentration of **1** at the start of the experiment, and θ is the fraction of thiolate at the experimental pH, calculated as $\theta = \frac{[\text{thiolate}]}{[\text{thiol}]} = \frac{1}{1 + 10^{\text{pK}_a - \text{pH}}}$. The rate constant was interpreted as the slope of the resulting linear plot. The reported rate constant is the average of three trials.

General Synthesis Procedures.

All reactions were performed in single-neck, oven-dried (130 °C for 12 h), round-bottomed flasks fitted with rubber septa under a positive pressure of argon. Organic solutions were concentrated by rotary evaporator at 30–50 °C, as noted. Analytical thin-layered chromatography (TLC) was performed using plates precoated with silica gel (Sigma-Aldrich; 60 Å, 17 µm particle size), impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by exposure to ultraviolet (UV) light. Flash-column chromatography was performed using a Biotage Isolera One automated flash purification system with Biotage SNAP KP-Sil cartridges.

Analysis of Synthetic Molecules.

¹H NMR spectra were recorded at 25 °C at 400 MHz. Chemical shifts are expressed in parts per million (ppm, δ scale) downfield from tetramethylsilane and are referenced to residual protium in the NMR (CDCl₃, δ 7.26; CD₃OD, δ 3.31). Data are represented as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet and/or multiple resonances, br = broad, app = apparent), coupling constant in Hertz, and integration. Proton-decoupled carbon nuclear magnetic resonance spectra (¹³C NMR) were recorded at 100 MHz at 25 °C. Chemical shifts are expressed in parts per million (ppm, δ scale) downfield from tetramethylsilane and are referenced to the carbon resonances of the solvent (CDCl₃, δ 77.0; CD₃OD, δ 49.0). Fluorine nuclear magnetic resonance spectra (¹⁹F NMR) were recorded at 376 MHz at 25 °C. Chemical shifts are expressed in parts per million (ppm, δ scale) and are referenced indirectly to the ¹H frequency of CHCl₃. High-resolution mass spectrometry (HRMS) data were obtained using a Waters Xevo-G2 XS qTOF instrument.

Sample Preparation for Mass Spectrometry Analysis of Reaction Sites of 1.

Full-length NC (1–55) or Gag p6 were incubated with 10-fold molar excess of **1** at 30 °C for 16 h with shaking. After the reaction, the unreacted **1** was removed by desalting on a Zeba desalting column (Thermo), and the samples were diluted into 8 M urea for digestion. For focused analysis on sites of lysine modification, the samples were first treated with 10 mM DTT at 37 °C for 1 h; this step was omitted for cysteine modification. Next, cysteines were alkylated with 50 mM iodoacetamide at room temperature for 30 min. Finally, the proteins were diluted with 50 mM ammonium bicarbonate to reduce the urea concentration to 1 M prior to digestion. NC was digested with 1:100 endoproteinase GluC or endoproteinase ArgC at 37 °C overnight; Gag p6 was digested with 1:100 trypsin or endoproteinase ArgC at 37 °C overnight. Digestion reactions were quenched by acidification with trifluoroacetic acid, and peptides were desalted on C18 spin columns (Thermo) and dried.

Gag isolated from viral particles was fractionated by reverse-phase HPLC. Individual fractions were digested with trypsin at 37 °C overnight following the filter aided sample preparation (FASP) method.³⁹ The peptides were desalted on C18 spin columns (Thermo) and dried. Alternately, for the analysis of Gag–Pol isolated from viral particles, lysed virions were separated by SDS-PAGE and individual gel slices subjected to in-gel trypsin digestion, as described.⁴⁰

For the analysis of the stoichiometry of Gag modification by **1**, Gag p6 was incubated with 10-fold molar excess of **1** at 30 °C for either 4 or 24 h with shaking and then the remaining **1** removed by passing the protein over a Zeba desalting column (Thermo). Isotopic labeling of the unmodified lysines in Gag p6 was performed as described.¹⁶ Briefly, after denaturation in urea, the reacted Gag p6 was reacted with ¹³C₂-acetic anhydride twice at 60 °C for 30 min. Following modification, the protein was digested with trypsin as described above.

Mass Spectrometry Analysis.

The dried peptides were solubilized in 5% acetonitrile, 0.5% acetic acid, water for mass spectrometry analysis. They were trapped on a trapping column and separated on a 75 μ m x 15 cm, 2 μ m Acclaim PepMap reverse-phase column (Thermo Scientific) using an UltiMate 3000 RSLCnano HPLC (Thermo Scientific). Peptides were separated at a flow rate of 300 nL/min followed by online analysis by tandem mass spectrometry using a Thermo Orbitrap Fusion mass spectrometer. Peptides were eluted into the mass spectrometer using a linear gradient from 96% mobile phase A (0.1% formic acid in water) to 55% mobile phase B (0.1% formic acid in acetonitrile). Parent full-scan mass spectra were collected in the Orbitrap mass analyzer set to acquire data at 120 000 fwhm resolution; ions were then isolated in the quadrupole mass filter, fragmented within the HCD cell (HCD normalized energy 32%, stepped $\pm$ 3%), and the product ions were analyzed in the ion trap. Proteome Discoverer 2.1 (Thermo) was used to search the data against all proteins from the UniProt database using SequestHT. Cysteine carbamidomethylation was set as a fixed modification for reactions focused on lysine modification; methionine oxidation and cysteine and/or lysine acetylation were set as a variable modifications. The precursor mass tolerance was 10 ppm, and the fragment mass tolerance was 0.6 Da. The maximum number of missed cleavages was set to 10, and the maximum number of equal modifications was also set to 10. The Percolator node was used to score and rank peptide and protein matches using a 1% false discovery rate.⁴¹ Acetylation site localization confidence was assessed using the PTM-RS node of Proteome Discoverer,⁴² and all modified spectra were manually validated. Quantitation of stoichiometry of acetylation was determined by measuring the intensity of the immonium ions in the MS/MS spectra, as described.¹⁶

HIV Gag NMR Sample Preparation and Spectroscopy.

All NMR samples were prepared in a buffer containing 20 mM sodium phosphate, pH 6.5, 300 mM NaCl, 0.1 mM ZnCl₂, 93% H₂O/7% D₂O (v/v), and 1 mM tris(2-carboxyethyl) phosphine (TCEP). Mercaptobenzamide **1** was dissolved in dimethyl sulfoxide (DMSO) to generate a 100 mM stock solution. The resulting solution was added to 300 μ L of NMR samples containing Gag p6 to achieve a final mercaptobenzamide **1** concentration of 0.5 mM. NMR titration experiments were performed using 50 μ M Gag p6 (in subunits) and 0.5 mM **1** (protein to compound molar ratio of ~1:10).

All heteronuclear NMR experiments were recorded at 30 °C on a Bruker 500 MHz spectrometer equipped with a z-gradient triple-resonance cryoprobe. NMR spectra were processed using NMRPipe⁴³ and analyzed using the CCPN software suite.⁴⁴ The backbone resonance assignments for Gag p6 were taken from our previously published work.^{19,34,45}

Preparation of 1 and Darunavir Solutions.

Stock solutions (100 mM in DMSO) of **1** and the active site clinical inhibitor darunavir (DRV) were diluted to 300 μ M in 5 mM sodium acetate, pH 6.

N-Terminal Autoprocessing.

A stock solution of the TFR-PR precursor (1.35 mg/mL, 39.2 μ M in dimer concentration) maintained in 12 mM HCl was diluted with 2.3 volumes of 5 mM sodium acetate buffer at pH 6 in the presence or absence of inhibitor, followed immediately by the addition of 3.3 volumes of 100 mM acetate buffer at pH 5. Aliquots (15 μ L) were drawn at 5, 10, and 30 min, mixed with 6 μ L of sample loading buffer, and 16 μ L of this mixture was subjected to SDS-PAGE on 4–20% gradient Mini-Protean TGX Precast Gel (Bio-Rad). Reactions were also carried out in a final concentration of 50 μ M SAMT-247 or 7.5 μ M DRV for 30 min, followed by SDS-PAGE.

Protease Assays.

Assays were carried out in 50 mM sodium acetate buffer (pH 5) at a final concentration of 200 nM mature protease either in the absence or presence of 20 μ M **1** and 380 μ M of the chromogenic substrate (Lys–Ala–Arg–Val–Nle–[4-nitrophenylalanine]–Glu–Ala–Nle–NH₂, California Peptide Research, Napa, CA⁴⁶) in a total volume of 120 μ L at 28 °C. Autoprocessing was initiated by folding the precursor TFR-PR either with or without **1** (50 μ M) as described above for monitoring the reaction by SDS-PAGE and assayed at a final concentration of 1 μ M precursor (as dimer). A positive control for inhibition of autoprocessing was carried out using 5 μ M of DRV.

Transmission Electron Microscopy and Western Blotting.

Micrographs of positive-stained virions were obtained as previously described.⁴⁷ Western blotting analyses were performed as previously described.¹³

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This research was funded by the Intramural Research Programs of the National Cancer Institute and the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health. Additional funding was from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. We thank Annie Aniana for technical assistance and Kunio Nagashima for electron microscopy. The following reagent was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: Darunavir (Cat# 11447) from Tibotec, Inc.

REFERENCES

- (1). Ngo-Giang-Huong N; Huynh THK; Dagnra AY; Toni TD; Maiga AI; Kania D; Eymard-Duvernay S; Peeters M; Soulie C; Peytavin G; et al. Prevalence of pretreatment HIV drug resistance in West African and Southeast Asian countries. *J. Antimicrob. Chemother.* 2019, 74 (2), 462–467. [PubMed: 30418575]

- (2). Chimukangara B; Kharsany ABM; Lessells RJ; Naidoo K; Rhee SY; Manasa J; Graf T; Lewis L; Cawood C; Khanyile D; Diallo K; Ayalew KA; Shafer RW; Hunt G; Pillay D; Abdool SK; de Oliveira T Moderate-to-High Levels of Pretreatment HIV Drug Resistance in KwaZulu-Natal Province, South Africa. *AIDS Res. Hum. Retroviruses* 2019, 35 (2), 129–138. [PubMed: 30430843]
- (3). Sundquist WI; Krausslich HG HIV-1 assembly, budding, and maturation. Cold Spring Harbor Perspect. Med 2012, 2 (8), a015420.
- (4). Tang C; Louis JM; Aniana A; Suh JY; Clore GM Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* 2008, 455 (7213), 693–6. [PubMed: 18833280]
- (5). Martin DE; Salzwedel K; Allaway GP Bevirimat: a novel maturation inhibitor for the treatment of HIV-1 infection. *Antivir. Chem. Chemother* 2008, 19 (3), 107–13. [PubMed: 19024627]
- (6). De Clercq E The nucleoside reverse transcriptase inhibitors, nonnucleoside reverse transcriptase inhibitors, and protease inhibitors in the treatment of HIV infections (AIDS). *Adv. Pharmacol* 2013, 67, 317–58. [PubMed: 23886005]
- (7). Gorelick RJ; Gagliardi TD; Bosche WJ; Wiltout TA; Coren LV; Chabot DJ; Lifson JD; Henderson LE; Arthur LO Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. *Virology* 1999, 256 (1), 92–104. [PubMed: 10087230]
- (8). Guo J; Wu T; Kane BF; Johnson DG; Henderson LE; Gorelick RJ; Levin JG Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol* 2002, 76 (9), 4370–8. [PubMed: 11932404]
- (9). Maynard AT; Covell DG Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *J. Am. Chem. Soc* 2001, 123 (6), 1047–58. [PubMed: 11456658]
- (10). Deshmukh L; Tugarinov V; Appella DH; Clore GM Targeting a Dark Excited State of HIV-1 Nucleocapsid by Antiretroviral Thioesters Revealed by NMR Spectroscopy. *Angew. Chem., Int. Ed* 2018, 57 (10), 2687–2691.
- (11). Miller Jenkins LM; Byrd JC; Hara T; Srivastava P; Mazur SJ; Stahl SJ; Inman JK; Appella E; Omichinski JG; Legault P Studies on the mechanism of inactivation of the HIV-1 nucleocapsid protein NCp7 with 2-mercaptobenzamide thioesters. *J. Med. Chem* 2005, 48 (8), 2847. [PubMed: 15828823]
- (12). Miller Jenkins LM; Hara T; Durell SR; Hayashi R; Inman JK; Piquemal JP; Gresh N; Appella E Specificity of acyl transfer from 2-mercaptobenzamide thioesters to the HIV-1 nucleocapsid protein. *J. Am. Chem. Soc* 2007, 129 (36), 11067. [PubMed: 17705474]
- (13). Miller Jenkins LM; Ott DE; Hayashi R; Coren LV; Wang D; Xu Q; Schito ML; Inman JK; Appella DH; Appella E Small-molecule inactivation of HIV-1 NCp7 by repetitive intracellular acyl transfer. *Nat. Chem. Biol* 2010, 6 (12), 887. [PubMed: 20953192]
- (14). Zhang Y; Mantravadi PK; Jobbagy S; Bao W; Koh JT Antagonizing the Androgen Receptor with a Biomimetic Acyltransferase. *ACS Chem. Biol.* 2016, 11 (10), 2797–2802. [PubMed: 27548116]
- (15). Hartman TL; Yang L; Helfrick AN; Hassink M; Shank NI; George Rosenker K; Scerba MT; Saha M; Hughes E; Wang AQ; Xu X; Gupta P; Buckheit RW Jr.; Appella DH Preclinical evaluation of a mercaptobenzamide and its prodrug for NCp7-targeted inhibition of human immunodeficiency virus. *Antiviral Res.* 2016, 134, 216–225. [PubMed: 27568924]
- (16). Fan J; Baeza J; Denu JM Investigating Histone Acetylation Stoichiometry and Turnover Rate. *Methods Enzymol.* 2016, 574, 125–148. [PubMed: 27423860]
- (17). Ott DE; Nigida SM Jr.; Henderson LE; Arthur LO The majority of cells are superinfected in a cloned cell line that produces high levels of human immunodeficiency virus type 1 strain MN. *J. Virol* 1995, 69 (4), 2443. [PubMed: 7884892]
- (18). Henderson LE; Bowers MA; Sowder RC 2nd; Serabyn SA; Johnson DG; Bess JW Jr.; Arthur LO; Bryant DK; Fenselau C Gag proteins of the highly replicative MN strain of human immunodeficiency virus type 1: posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol* 1992, 66 (4), 1856. [PubMed: 1548743]

- (19). Deshmukh L; Ghirlando R; Clore GM Conformation and dynamics of the Gag polypeptide of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A* 2015, 112 (11), 3374–9. [PubMed: 25713345]
- (20). Ott DE; Coren LV; Chertova EN; Gagliardi TD; Nagashima K; Sowder RC 2nd; Poon DT; Gorelick RJ Elimination of protease activity restores efficient virion production to a human immunodeficiency virus type 1 nucleocapsid deletion mutant. *J. Virol* 2003, 77 (10), 5547–56. [PubMed: 12719547]
- (21). Accola MA; Hoglund S; Gottlinger HG A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J. Virol* 1998, 72 (3), 2072. [PubMed: 9499062]
- (22). Somsouk M; Dunham RM; Cohen M; Albright R; Abdel-Mohsen M; Liegler T; Lifson J; Piatak M; Gorelick R; Huang Y; Wu Y; Hsue PY; Martin JN; Deeks SG; McCune JM; Hunt PW The immunologic effects of mesalamine in treated HIV-infected individuals with incomplete CD4+ T cell recovery: a randomized crossover trial. *PLoS One* 2014, 9 (12), No. e116306. [PubMed: 25545673]
- (23). Shehu-Xhilaga M; Crowe SM; Mak J Maintenance of the Gag/Gag–Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol* 2001, 75 (4), 1834–41. [PubMed: 11160682]
- (24). Wang D; Lu W; Li F Pharmacological intervention of HIV-1 maturation. *Acta Pharm. Sin. B* 2015, 5 (6), 493–9. [PubMed: 26713265]
- (25). Perrier M; Bertine M; Le Hingrat Q; Joly V; Visseaux B; Collin G; Landman R; Yazdanpanah Y; Descamps D; Charpentier C Prevalence of gag mutations associated with in vitro resistance to capsid inhibitor GS-CA1 in HIV-1 antiretroviral-naïve patients. *J. Antimicrob. Chemother.* 2017, 72 (10), 2954–2955. [PubMed: 29091184]
- (26). Schito ML; Goel A; Song Y; Inman JK; Fattah RJ; Rice WG; Turpin JA; Sher A; Appella E In vivo antiviral activity of novel human immunodeficiency virus type 1 nucleocapsid p7 zinc finger inhibitors in a transgenic murine model. *AIDS Res. Hum. Retroviruses* 2003, 19 (2), 91–101. [PubMed: 12639244]
- (27). Schito ML; Soloff AC; Slovit D; Trichel A; Inman JK; Appella E; Turpin JA; Barratt-Boyes SM Preclinical evaluation of a zinc finger inhibitor targeting lentivirus nucleocapsid protein in SIV-infected monkeys. *Curr. HIV Res.* 2006, 4 (3), 379–86. [PubMed: 16842089]
- (28). Wallace GS; Cheng-Mayer C; Schito ML; Fletcher P; Miller Jenkins LM; Hayashi R; Neurath AR; Appella E; Shattock RJ Human immunodeficiency virus type 1 nucleocapsid inhibitors impede trans infection in cellular and explant models and protect nonhuman primates from infection. *J. Virol* 2009, 83 (18), 9175–82. [PubMed: 19587055]
- (29). Yang Y; Zhu J; Hassink M; Jenkins LMM; Wan Y; Appella DH; Xu J; Appella E; Zhang X A novel preventive strategy against HIV-1 infection: combinatorial use of inhibitors targeting the nucleocapsid and fusion proteins. *Emerging Microbes Infect.* 2017, 6 (1), No. 1.
- (30). Jenkins LM; Durell SR; Maynard AT; Stahl SJ; Inman JK; Appella E; Legault P; Omichinski JG Comparison of the specificity of interaction of cellular and viral zinc-binding domains with 2-mercaptobenzamide thioesters. *J. Am. Chem. Soc.* 2006, 128 (36), 11964–76. [PubMed: 16953638]
- (31). Lee SK; Potempa M; Swanstrom R The choreography of HIV-1 proteolytic processing and virion assembly. *J. Biol. Chem.* 2012, 287 (49), 40867–74. [PubMed: 23043111]
- (32). Turpin JA; Schito ML; Miller Jenkins L; Inman JK; Appella E Topical microbicides: a promising approach for controlling the AIDS pandemic via retroviral zinc finger inhibitors. *Adv. Pharmacol.* 2008, 56, 229. [PubMed: 18086414]
- (33). Nikolayevskiy H; Scerba MT; Deschamps JR; Appella DH Reaction Kinetics Direct a Rational Synthesis of an HIV-1 Inactivator of Nucleocapsid Protein 7 and Provide Mechanistic Insight into Cellular Metabolism and Antiviral Activity. *Chem. - Eur. J* 2018, 24 (38), 9485–9489. [PubMed: 29653024]
- (34). Deshmukh L; Louis JM; Ghirlando R; Clore GM Transient HIV-1 Gag-protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. *Proc. Natl. Acad. Sci. U. S. A.* 2016, 113 (44), 12456–12461. [PubMed: 27791180]

- (35). Louis JM; Aniana A; Weber IT; Sayer JM Inhibition of autoprocessing of natural variants and multidrug resistant mutant precursors of HIV-1 protease by clinical inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 108 (22), 9072–7. [PubMed: 21576495]
- (36). Louis JM; Clore GM; Gronenborn AM Autoprocessing of HIV-1 protease is tightly coupled to protein folding. *Nat. Struct. Biol* 1999, 6 (9), 868–75. [PubMed: 10467100]
- (37). Louis JM; Ishima R; Aniana A; Sayer JM Revealing the dimer dissociation and existence of a folded monomer of the mature HIV-2 protease. *Protein Sci.* 2009, 18 (12), 2442–53. [PubMed: 19798742]
- (38). Sayer JM; Agniswamy J; Weber IT; Louis JM Autocatalytic maturation, physical/chemical properties, and crystal structure of group N HIV-1 protease: relevance to drug resistance. *Protein Sci.* 2010, 19 (11), 2055–72. [PubMed: 20737578]
- (39). Wisniewski JR; Zougman A; Nagaraj N; Mann M Universal sample preparation method for proteome analysis. *Nat. Methods* 2009, 6 (5), 359–62. [PubMed: 19377485]
- (40). Shevchenko A; Tomas H; Havli J; Olsen JV; Mann M In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 2006, 1 (6), 2856. [PubMed: 17406544]
- (41). Kall L; Canterbury JD; Weston J; Noble WS; MacCoss MJ Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 2007, 4 (11), 923–5. [PubMed: 17952086]
- (42). Taus T; Kocher T; Pichler P; Paschke C; Schmidt A; Henrich C; Mechtler K Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* 2011, 10 (12), 5354–62. [PubMed: 22073976]
- (43). Delaglio F; Grzesiek S; Vuister GW; Zhu G; Pfeifer J; Bax A NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* 1995, 6 (3), 277. [PubMed: 8520220]
- (44). Vranken WF; Boucher W; Stevens TJ; Fogh RH; Pajon A; Llinas M; Ulrich EL; Markley JL; Ionides J; Laue ED The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Struct., Funct., Genet* 2005, 59 (4), 687–96. [PubMed: 15815974]
- (45). Deshmukh L; Tugarinov V; Louis JM; Clore GM Binding kinetics and substrate selectivity in HIV-1 protease-Gag interactions probed at atomic resolution by chemical exchange NMR. *Proc. Natl. Acad. Sci. U. S. A.* 2017, 114 (46), E9855–E9862. [PubMed: 29087351]
- (46). Richards AD; Phylip LH; Farmerie WG; Scarborough PE; Alvarez A; Dunn BM; Hirel PH; Konvalinka J; Strop P; Pavlickova L; et al. Sensitive, soluble chromogenic substrates for HIV-1 proteinase. *J. Biol. Chem.* 1990, 265 (14), 7733. [PubMed: 2186027]
- (47). Gonda MA; Aaronson SA; Ellmore N; Zeve VH; Nagashima K Ultrastructural studies of surface features of human normal and tumor cells in tissue culture by scanning and transmission electron microscopy. *J. Natl. Cancer Inst.* 1976, 56 (2), 245–63. [PubMed: 1255758]

Figure 1.

Zinc-finger inhibitor SAMT-247 (**1**) modifies and unfolds the NC domain of Gag. (A) Schematic representation of the proposed reaction mechanism for **1**, including relevant chemical structures. The C-terminal zinc-coordinated knuckle of NC is in rapid equilibrium with a small population of an unligated form that has a free thiol.¹⁰ In the presence of **1**, this free thiol can be acetylated and the acetyl group subsequently transferred to a nearby amino group on a lysine side chain. During this process, the coordination to zinc is disrupted and the protein unfolds.¹⁰ (B) Lysine side chains of NC that are acetylated after treatment with **1**. Peptide sequence numbers are shown in black. "Ac" represents an acetyl group attached to the side chain amine of lysine as observed by mass spectrometry. The font size reflects the relative number of peptide spectral matches (PSMs), and the number next to the "Ac" in parentheses is the actual PSM number as a result of modification by **1**.

Figure 2.

NMR analysis of Gag unfolding following incubation with **1**. (A) Time-course analysis of $^1\text{H}/^{15}\text{N}$ cross-peak height ratios for Gag p6 upon addition of **1**. The decrease in cross-peak intensities is first manifested for residues within the C-terminal zinc knuckle of the NC domain, which are highlighted by the transparent blue bar. Secondary structure elements and Gag domain organization are indicated above the panel. Note that $^1\text{H}/^{15}\text{N}$ cross-peaks for several regions within CA are not observed in the $^1\text{H}-^{15}\text{N}$ TROSY correlation spectrum of Gag p6 as a consequence of intermolecular interactions involving the C-terminal dimerization domain (CTD) of CA. (B) Time course of the cumulative normalized cross-peak heights (circles) and volumes (lines) after addition of **1**. Values are shown for all domains of Gag p6, with MA in gold, CA-NTD in blue, CA-CTD in red, SP1 in green, N-terminal knuckle of NC in black, and C-terminal knuckle of NC in pink. The decreases in cross-peak heights (circles) and volumes (lines) are completely superimposable within experimental errors, indicating that the decrease in intensity can be attributed entirely to the irreversible formation of large, NMR invisible, aggregates and not to an exchange process in the intermediate exchange regime on the chemical shift time scale. The latter would result in a decrease in peak heights while leaving the peak volumes unaffected. Aggregation is initiated by Zn^{2+} ejection within the C-terminal zinc knuckle of NC induced by compound **1**, which presumably results in unfolding and subsequent aggregation, potentially involving the formation of interchain disulfide linkages.

Identical Conditions ($\text{pH} = 7.4 \pm 0.3$, $T = 25^\circ\text{C}$, $[\text{SAMT-247 (1)}] = [\text{RSH}] = 1.5\text{ mM}$)

Figure 3.

(A) Intermolecular thioester–thiol exchange between **1** and thiols cysteine (**3a**), homocysteine (**4a**), the Cys–Gly dipeptide **5a**, or the Cys– β -Ala dipeptide **6a**, followed by intramolecular thioester–amine exchange under *identical* experimental conditions (93:7 PBS 10 $\times$ /D₂O, $\text{pH} = 7.4 \pm 0.3$, 25°C , $[\text{SAMT-247(1)}] = [\text{thiol}] = 1.5\text{ mM}$). Similar rates of intermolecular thioester–thiol exchange but divergent rates of intramolecular thioester–amine exchange led to differences in the observed accumulation of the peptide thioester intermediate (**3b**, **4b**, **5b**, or **6b**). (B) Second order thioester–thiol exchange rate constants ($\text{M}^{-1}\text{ s}^{-1}$). Reported data is the average of three trials. (C) Thioester (**3b**, **4b**, **5b**, or **6b**)

concentration (μM) after intermolecular thioester–thiol exchange with SAMT-247 (**1**) as a function of time (s).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Residues that are targets of reaction with **1** are surface exposed on Gag. (A) Model of Gag p6¹⁹ showing the positions of residues found to be modified following treatment with **1**, with focuses on the modifications in MA (B), CA N-terminal domain (NTD) (C), and CA C-terminal domain (CTD) (D). Protein domains are shown in ribbon format, with MA in cyan, CA in pink, and NC in green. The modified Cys side chains are shown as cyan sticks and modified Lys side chain as green sticks. In NC, the zinc-coordinating residues and zinc are shown as gray.

Figure 5.

Functional consequences of treatment of HIV_{MN}-infected Clone 4 H9 cells with **1**. (A,B) Transmission electron micrographs of the morphology of virions produced in the absence and presence of **1**. (A) Panels under the “untreated” label show combined fields of mature virus produced in the absence of **1**. The panel under the “**1**-treated” label shows a single field of virions produced in the presence of **1** with arrows (labeled 1 and 2) showing the two observed immature virus morphologies that are described in the text. These morphologies are similar to a protease mutant (arrow 1) or an SP1 mutant (arrow 2). (B) Examples of budding structures and virion close-up images. A 100 nm reference bar is provided for reference. Panels under the “untreated” label show mature virus. Panels under the “**1**-treated” label show close-up images of the two immature virus morphologies with the numbers 1 and 2 corresponding to the arrows in Figure 5A. (C,D) Immunoblot analysis of virions produced from untreated (–) and treated (+) cells as separated on reducing (C) and nonreducing (D) gels. Proteins were separated on reducing gels. The FL₃₀₂ lane is a gel with proteins visualized by tryptophan-FRET fluorescence by 302 nm light, and subsequent lanes are immunoblots of the same gel. Antibodies and antisera used are denoted above their respective lanes with molecular marker positions provided on the left and the identification of the bands presented on the right. The αRT-CT antibody is specific for the C-terminal portion of p66^{RT}, which is missing from the p51^{RT} cleavage product. In (D), the expected positions of the reduced proteins are indicated at right, and a lighter exposure of the nondenatured RT-CT blot is present at right to better see the protein complexes. (E) Schematic model of the proposed mechanism, by which treatment with **1**

blocks PR function. Reaction of **1** with Cys residues in Gag leads to both modification of proximal Lys residues and loss of zinc coordination in NC; both cysteine and lysine modification likely destabilize zinc coordination. The Cys residues are thereby available to form intermolecular disulfide bonds with neighboring Gag and Gag–Pol molecules in the high local concentration environment of the budding virion. As the quantity of Gag–Pol is low compared to that of Gag, it is likely that most Gag–Pol will form intermolecular disulfide cross-links with Gag molecules (rather than other Gag–Pol molecules), which prevents the Gag–Pol dimerization required for autocatalytic activation of PR.

Table 1.

Time-Course Analysis of Cysteine Modification of Gag

annotated sequence	modification	peptide quantitation ^a	
		4 h	24 h
NWMTETLLVQNANPDCCKTILKALGPAATLEEMMTACQGVGGPGHK	CA C330	3.88×10^{05}	1.57×10^{06}
NWMTETLLVQNANPDCCKTILKALGPAATLEEMMTACQGVGGPGHK	CA C350	9.43×10^{04}	7.57×10^{05}
IVKCFNCGKEGHTARNCR	NC C395, C405	5.06×10^{04}	1.03×10^{06}
NQRKIVKCFNCGKEGHTARNCRAPRKKGWCKGK	NC K388, K391, C413, C416	9.51×10^{03}	5.40×10^{04}

^aPeptide abundance measured from the peak area for each peptide from the extracted ion chromatogram.

Table 2.

Time-Course Analysis of the Stoichiometry of Gag Lysine Modification

annotated sequence	modification	% reacted 4 h	% reacted 24 h
ASVLSGGELDRWEKIR	MA K18	<1	<1
EKIRLRPGGKKKY	MA K18		23
EKIRLRPGGKKKY	MA K27/K28	<1	45
KLKHIVW	MA K30, K32	9	20
NAWVKVVEEKAF	CA K152	<1	
NAWVKVVEEKAF	CA K157	<1	<1
VKVVEEKAF	CA K157, K162	3	
VKVVEEKAFSPVIMFSAI	CA K157, K162	<1	
GLNKIVRMV	CA K272		<1
DIRQGPKEPF	CA K290	<1	20
DIRQGPKEPFRDY	CA K290	<1	62
TSILDIRQGPKEPF	CA K290	<1	
KTILRAEQASQEVKNW	CA K314	<1	10
RAEQASQEVKNW	CA K314	<1	7
KALGPAATL	CA K335		13
KALGPAATLEEMMTACQGVGGPGHKARVL	CA K335, K359		<1
RNQRKIVKCF	NC K388, K391	6	14
KIVKCFNGGKEGHTAR	NC K388, K391, K397		16
KIVKCFNGGKEGHTAR	NC K388/K391, K397	<1	

Table 3.

Gag Lysine Reaction Sites Observed in Vitro and in Virions from Treated Cells

site of modification ^a	gag domain	experiment type ^b		
Lys18	MA	in vitro	in vivo	
Lys38	MA	in vitro		
Lys40	MA	in vitro		
Lys42	MA	in vitro		
Lys76	MA	(R76)	in vivo	basal
Lys91	MA	(R91)	in vivo	
Lys95	MA	in vitro	in vivo	
Lys98	MA	in vitro	in vivo	
Lys103	MA	in vitro	in vivo	basal
Lys110	MA	in vitro	in vivo	
Lys112	MA	in vitro	in vivo	
Lys113	MA	in vitro	in vivo	basal
Lys114	MA	in vitro		
Lys157 (Lys160)	CA	in vitro	in vivo	
Lys272 (Lys275)	CA	in vitro		
Lys290 (Lys293)	CA		in vivo	basal
Lys314 (Lys317)	CA		in vivo	
Lys331 (Lys334)	CA		in vivo	
Lys359 (Lys362)	CA	in vitro	in vivo	
Lys388 (Lys391)	NC	in vitro		
Lys391 (Lys394)	NC	in vitro		
Lys397 (Lys400)	NC	in vitro	in vivo	
Lys406	NC	(R403)	in vivo	
Lys410 (Lys413)	NC		in vivo	
Lys415 (Lys418)	NC	in vitro		
Lys418 (Lys421)	NC	in vitro		
Lys424 (Lys427)	NC	in vitro		
Lys475 (Lys478)	p6	NP ^c	in vivo	
(Lys481)	p6	NP	in vivo	
Lys481 (Lys482)	p6	NP	in vivo	
Arg490 (Lys488)	p6	NP	in vivo	basal

^aNumbering is for HIV-1HXB2, as used in the in vitro experiments; numbering in parentheses is for HIV-1MN, as used in the cellular experiments.

^bNumbering for arginine residues indicates that the modified lysine in HIV-1MN is an arginine in HIV-1HXB2.

^cNP = not present.