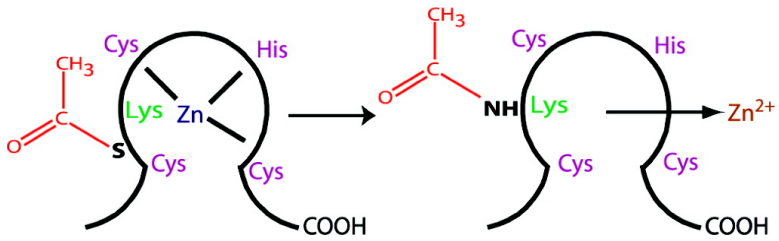
Article

Specificity of Acyl Transfer from 2-Mercaptobenzamide Thioesters to the HIV-1 Nucleocapsid Protein

Lisa M. Miller Jenkins, Toshiaki Hara, Stewart R. Durell, Ryo Hayashi, John K. Inman, Jean-Philip Piquemal, Nohad Gresh, and Ettore Appella

J. Am. Chem. Soc., **2007**, 129 (36), 11067-11078• DOI: 10.1021/ja071254o • Publication Date (Web): 18 August 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/18/2007

Specificity of Acyl Transfer from 2-Mercaptobenzamide Thioesters to the HIV-1 Nucleocapsid Protein

Lisa M. Miller Jenkins,[†] Toshiaki Hara,[†] Stewart R. Durell,[†] Ryo Hayashi,[†] John K. Inman,[‡] Jean-Philip Piquemal,[§] Nohad Gresh,[⊥] and Ettore Appella*[,†]

Contribution from the Laboratory of Cell Biology, NCI, and Laboratory of Immunology, NIAID, National Institutes of Health, Bethesda, Maryland 20892, Laboratoire de Chimie Théorique, Université Pierre-et-Marie-Curie (Paris VI), UMR 7616 CNRS, CC137, 4 Place Jussieu, 75005 Paris, France, and Laboratoire de Pharmacochimie Moléculaire et Cellulaire, U648 INSERM, IFR Biomédicale, 45, rue des Saints-Pères, 75006 Paris, France

Received February 21, 2007; E-mail: appellae@pop.nci.nih.gov

Abstract: The HIV-1 nucleocapsid protein (NCp7) is a small, highly conserved protein with two zinc-binding domains that are essential for the protein's function. Molecules that bind to and inactivate NCp7 are currently being evaluated as new antiviral drugs. In particular, derivatives based on a 2-mercaptobenzamide thioester template have been shown to specifically eject zinc from the C-terminal zinc-binding domain (ZD2) of NCp7 via acyl transfer from the thioester to a cysteine sulfur. In this study, mutational analysis of the NCp7 amino acid sequence has been used to investigate the specificity of the interaction between ZD2 and a 2-mercaptobenzamide thioester compound using UV−vis spectroscopy and mass spectrometry to monitor the rate of metal ejection from NCp7 mutant peptides and sites of acylation, respectively. We were able to extend the previously reported mechanism of action of these thioester compounds to include a secondary S to N intramolecular acyl transfer that occurs after the primary acyl transfer from the thioester to a cysteine side chain in the protein. Structural models of the thioester/ZD2 complex were then examined to identify the most likely binding orientation. We determined that position $x + 1$ (where x is Cys_{36}) needs to be an aromatic residue for reactivity and a hydrogen-bond donor in position $x + 9$ is important for optimal reactivity. A basic residue (lysine or arginine) is required at position $x + 2$ for the correct fold, while a lysine residue is needed for reactivity involving S to N acyl transfer. We report highly specific interactions between 2-mercaptobenzamide thioester compounds and NCp7 that offer a structural basis for refining and designing new antiretroviral therapeutics, directed toward a target that is resistant to viral mutation.

Introduction

The HIV-1 nucleocapsid protein (NCp7) is widely regarded as a potential target for the development of antiviral drugs. Traditional combination therapies can significantly reduce viral replication, but they cannot completely eradicate the virus. Therefore, long-term treatment is necessary to maintain low viral levels. Unfortunately, current treatments are often complicated by problems with patient compliance and metabolic side effects.[1−3] Over the long-term, strains of the virus emerge that are resistant to the antiviral therapies.[4] Therefore, it is essential to develop new antiviral drugs aimed at novel targets that are not prone to mutation.

NCp7 is a small, highly conserved, multifunctional protein that contains two zinc-binding domains. In each zinc-binding domain, three cysteine residues and one histidine residue coordinate one zinc ion, forming the $CysX_2CysX_4HisX_4Cys$ motif (where X is any amino acid) that is unique to retroviruses. The zinc-binding domains define the principal structural elements in the protein and have been shown to form two tight reverse turns.[5−8] NMR studies have revealed that the two zinc-binding domains can transiently interact and, thus, do not form completely independent structural units.[5,9] NCp7 functions at many stages in the viral lifecycle, including reverse transcription, integration, virion assembly, and RNA packaging.[10−14] In these

[†] NCI, National Institutes of Health.
[‡] NIAID, National Institutes of Health.
[§] Université Pierre-et-Marie-Curie.
[⊥] U648 INSERM, IFR Biomédicale.

(1) Cohen, J. *Science* **1997**, *277*, 32−33.
(2) Finzi, D.; Hermankova, M.; Pierson, T.; Carruth, L. M.; Buck, C.; Chaisson, R. E.; Quinn, T. C.; Chadwick, K.; Margolick, J.; Brookmeyer, R.; Gallant, J.; Markowitz, M.; Ho, D. D.; Richman, D. D.; Siliciano, R. F. *Science* **1997**, *278*, 1295−1300.
(3) Wong, J. K.; Hezareh, M.; Gunthard, H. F.; Havlir, D. V.; Ignacio, C. C.; Spina, C. A.; Richman, D. D. *Science* **1997**, *278*, 1291−1295.
(4) Tozser, J. *Ann. N.Y. Acad. Sci.* **2001**, *946*, 15.
(5) Morellet, N.; Jullian, N.; De Rocquigny, H.; Maigret, B.; Darlix, J.; Roques, B. P. *EMBO J.* **1992**, *11*, 3059−3065.
(6) Omichinski, J. G.; Clore, G. M.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M. *FEBS Lett.* **1991**, *292*, 25−30.
(7) South, T. L.; Blake, P. R.; Hare, D. R.; Summers, M. F. *Biochemistry* **1991**, *30*, 6342−6349.
(8) Summers, M. F.; South, T. L.; Kim, B.; Hare, D. R. *Biochemistry* **1990**, *29*, 329−340.
(9) Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F. *J. Mol. Biol.* **1998**, *279*, 633−649.
(10) Carteau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H.; Darlix, J. L.; Roques, B. P.; Kas, E.; Auclair, C. *J. Virol.* **1997**, *71*, 6225−6229.

10.1021/ja071254o CCC: $37.00

stages, NCp7 acts primarily as a nucleic acid-binding protein, interacting with both RNA and DNA.[10−14] The zinc-binding domains and several basic amino acids at the amino terminus of NCp7 have been shown to be required for sequence-specific RNA binding.[15,16] The crucial roles played by NCp7 in multiple stages of the HIV-1 lifecycle suggest that inhibiting its function would be particularly effective in halting the infectivity and progression of the virus. In addition, the primary structure of the protein is highly conserved and mutationally nonpermissive, indicating that drug-resistant mutations, common to other retroviral targets, would be slow to arise in NCp7.[17−20]

Several classes of compounds have been designed to inhibit NCp7. These compounds act primarily by covalently modifying zinc-coordinating cysteines, resulting in ejection of zinc and loss of protein function. Initial studies examined molecules based on 3-nitrosobenzamide (NOBA) which were highly active against NCp7 but also reacted nonspecifically with cellular zinc-binding proteins.[21,22] Further studies led to the development of compounds with greater selectivity, such as the disulfide benzamide (DIBA) and azodicarbonamide (ADA) compounds.[23,24] DIBA compounds were highly active against NCp7, both in vitro and in cell culture experiments.[25−32] It has been proposed that the DIBA compounds function via an oxidative mechanism in which the sulfur atom from one of the zinc-coordinated cysteine residues slowly reacts with the disulfide moiety of the DIBA compound to form a mixed disulfide bond.[28] A similar mechanism has been proposed for ADA compounds.[33]

More recently, molecules based on a mercaptobenzamide thioester have been developed to inactivate HIV-1 by targeting NCp7. These compounds are more water-soluble than the DIBA compounds and eject coordinated zinc from NCp7.[34,35] In addition, they have high antiviral activity with low cytotoxicity in cell culture and animal models.[36,37] Recently, we have examined the mechanism of action of these thioester compounds with NCp7.[38] We found that they preferentially interact with the second, C-terminal zinc-binding domain (ZD2) over the first, N-terminal zinc-binding domain (ZD1).[38] In our proposed mechanism, the thiol group of a zinc-coordinated cysteine ligand nucleophilically attacks the carbonyl carbon of the thioester compound resulting in covalent modification of the cysteine sulfur via an acyl transfer mechanism. This initial acylation weakens zinc coordination at the site of reaction. Subsequently, more acyl transfer reactions occur with other cysteine residues in the domain, which results in metal ejection, protein unfolding, and loss of function. Thus, the 2-mercaptobenzamide thioesters inactivate NCp7 by a mechanism different from previous inhibitors.

This paper details the studies we undertook to understand the reason for the preference of mercaptobenzamide thioesters for the C-terminal zinc-binding domain. As the amino acid spacing is identical and structure of the two zinc-binding domains of NCp7 are very similar, we hypothesized that the difference in reactivity between the two is determined by amino acids proximal to the zinc-coordinating residues, which would act to bind and favorably orient the thioester compounds for reaction with one or more cysteines of ZD2. On the basis of a comparison of the primary structure of ZD1 and ZD2, amino acid substitutions in NCp7 were chosen to determine which residues contribute to the reaction-site specificity. As in our previous studies, UV−vis spectroscopy of cobalt-substituted NCp7 was used to monitor the reaction rate, and mass spectrometry was employed to identify the sites of acetylation. These initial results were then used to guide three-dimensional (3-D) structural models of the possible docking orientations of the thioester on ZD2. In addition to the standard, classical physics-based force fields, the models were refined with the SIBFA polarizable molecular mechanics procedure, which has been shown to accurately reproduce the results from ab initio quantum-chemical computations in several model complexes extracted from ligand−zinc−metalloprotein simulations (for recent examples, see refs 39−41). This provided more refined

(11) Darlix, J. L.; Gabus, C.; Nugeyre, M. T.; Clavel, F.; Barre-Sinoussi, F. *J. Mol. Biol.* **1990**, *216*, 689−699.
(12) Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G. *J. Virol.* **1997**, *71*, 5178−5188.
(13) Negroni, M.; Buc, H. *Annu. Rev. Genet.* **2001**, *35*, 275−302.
(14) Sandefur, S.; Smith, R. M.; Varthakavi, V.; Spearman, P. *J. Virol.* **2000**, *74*, 7238−7249.
(15) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F. *J. Mol. Biol.* **2000**, *301*, 491−511.
(16) De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. *Science* **1998**, *279*, 384−388.
(17) Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Gottlinger, H. G. *J. Virol.* **1993**, *67*, 6159−6169.
(18) Housset, V.; De Rocquigny, H.; Roques, B. P.; Darlix, J. L. *J. Virol.* **1993**, *67*, 2537−2545.
(19) Poon, D. T.; Wu, J.; Aldovini, A. *J. Virol.* **1996**, *70*, 6607−6616.
(20) Tanchou, V.; Decimo, D.; Pechoux, C.; Lener, D.; Rogemond, V.; Berthoux, L.; Ottmann, M.; Darlix, J. L. *J. Virol.* **1998**, *72*, 4442−4447.
(21) Rice, W. G.; Schaeffer, C. A.; Graham, L.; Bu, M.; McDougal, J. S.; Orloff, S. L.; Villinger, F.; Young, M.; Oroszlan, S.; Fesen, M. R.; Pommier, Y.; Mendeleyev, J.; Kun, E. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9721−9724.
(22) Rice, W. G.; Schaeffer, C. A.; Harten, B.; Villinger, F.; South, T. L.; Summers, M. F.; Henderson, L. E.; Bess, J. W., Jr.; Arthur, L. O.; McDougal, J. S.; Orloff, S. L.; Mendeleyev, J.; Kun, E. *Nature* **1993**, *361*, 473−475.
(23) Rice, W. G.; et al. *Science* **1995**, *270*, 1194−1197.
(24) Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr.; Covell, D. G.; Wallqvist, A.; McDonnell, N. B.; DeGuzman, R. N.; Summers, M. F.; Zalkow, L.; Bader, J. P.; Haugwitz, R. D.; Sausville, E. A. *Nat. Med.* **1997**, *3*, 341−345.
(25) Domagala, J. M.; Bader, J. P.; Gogliotti, R. D.; Sanchez, J. P.; Stier, M. A.; Song, Y.; Vara Prasad, J. V. N.; Tummino, P. J.; Scholten, J.; Harvey, P.; Holler, T.; Gracheck, S.; Hupe, D.; Rice, W. G.; Schultz, R. *Bioorg. Med. Chem.* **1997**, *5*, 569−579.
(26) Domagala, J. M.; et al. *Drug Des. Discovery* **1997**, *15*, 49−61.
(27) Huang, M.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 1371−1381.
(28) Loo, J. A.; Holler, T. P.; Sanchez, J.; Gogliotti, R.; Maloney, L.; Reily, M. D. *J. Med. Chem.* **1996**, *39*, 4313−4320.
(29) Vara Prasad, J. V. N.; Loo, J. A.; Boyer, F. E.; Stier, M. A.; Gogliotti, R. D.; Turner, W. J.; Harvey, P. J.; Kramer, M. R.; Mack, D. P.; Scholten, J. D.; Gracheck, S. J.; Domagala, J. M. *Bioorg. Med. Chem.* **1998**, *6*, 1707−1730.
(30) Tummino, P. J.; Harvey, P. J.; McQuade, T.; Domagala, J.; Gogliotti, R.; Sanchez, J.; Song, Y.; Hupe, D. *Antimicrob. Agents Chemother.* **1997**, *41*, 394−400.
(31) Tummino, P. J.; Scholten, J. D.; Harvey, P. J.; Holler, T. P.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 969−973.
(32) Turpin, J. A.; Terpening, S. J.; Schaeffer, C. A.; Yu, G.; Glover, C. J.; Felsted, R. L.; Sausville, E. A.; Rice, W. G. *J. Virol.* **1996**, *70*, 6180−6189.
(33) Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. I.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 11341−11350.
(34) Goel, A.; Mazur, S. J.; Fattah, R. J.; Hartman, T. L.; Turpin, J. A.; Huang, M.; Rice, W. G.; Appella, E.; Inman, J. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 767−770.
(35) Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A.; Maynard, A.; Covell, D. G.; Rice, W. G.; Appella, E. *J. Med. Chem.* **1999**, *42*, 67−86.
(36) Schito, M. L.; Goel, A.; Song, Y.; Inman, J. K.; Fattah, R. J.; Rice, W. G.; Turpin, J. A.; Sher, A.; Appella, E. *AIDS Res. Hum. Retroviruses* **2003**, *19*, 91−101.
(37) Song, Y.; Goel, A.; Basrur, V.; Roberts, P. E.; Mikovits, J. A.; Inman, J. K.; Turpin, J. A.; Rice, W. G.; Appella, E. *Bioorg. Med. Chem.* **2002**, *10*, 1263−1273.
(38) Jenkins, L. M. M.; Byrd, J. C.; Hara, T.; Srivastava, P.; Mazur, S.; Stahl, S. J.; Inman, J. K.; Appella, E.; Omichinski, J. G.; Legault, P. *J. Med. Chem.* **2005**, *48*, 2847−2858.
(39) Antony, J.; Piquemal, J.-P.; Gresh, N. *J. Comput. Chem.* **2005**, *11*, 1131−1147.
(40) Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113−1130.

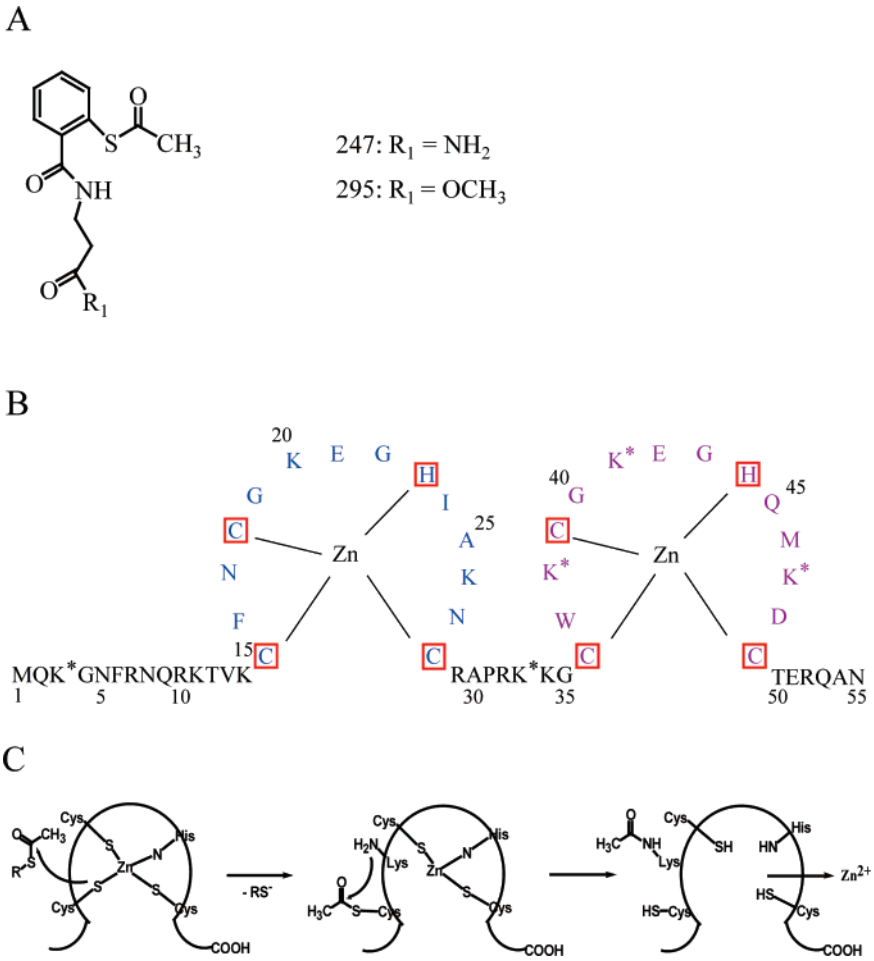


Figure 1. (A) Chemical structure of thioester compounds used in this study. (B) Primary sequence of NCp7. The zinc-coordinating residues are marked with red boxes. The residues in the N-terminal zinc-binding domain are blue and those of the C-terminal zinc-binding domain are purple. Lysine residues where acetyl transfer is observed are marked with an asterisk. (C) Proposed reaction mechanism for the 2-mercaptobenzamide thioester compounds showing the sequential acylation of cysteine and then lysine residues in ZD2 leading to zinc ejection.

estimates of the binding energies and orientations for the interacting functional groups. Finally, additional reaction rate experiments were conducted to identify the most likely binding orientation and, thus, further our understanding of the reaction mechanism of the mercaptobenzamide thioesters.

Experimental Methods

Sample Preparation. Compound 247 (*N*-[2-(acetylthio)benzoyl]-*β*-alaninamide) and compound 295 (*N*-[2-(acetylthio)benzoyl]-*β*-alanine methyl ester) (Figure 1A) were synthesized as previously described.[42] Further information on the chemical identification of compound 247 is reported in the Supporting Information. Full-length NCp7 (Figure 1B), $NCp7_{32-55}$, and related mutant peptides were chemically synthesized as previously described.[43] The synthesized peptides were purified on a C18 reversed-phase, high-performance liquid chromatography (RP-HPLC) column (Thomson Liquid Chromatography, VA), and the mass was verified by MALDI-TOF mass spectrometry. Purified NCp7 and $NCp7_{32-55}$ were refolded using a previously described procedure.[38] Briefly, lyophilized protein was first resuspended in 0.05% TFA with 5 molar equiv of $ZnCl_2$ (zinc-refolded NCp7) or $CoCl_2$ (cobalt-refolded NCp7). The solution was titrated to pH 6.0, and excess metal was precipitated by the addition of sodium phosphate pH 7.0. Zinc-refolded samples were lyophilized prior to use, while cobalt-refolded samples were used immediately.

UV−Vis Spectroscopy. Cobalt-refolded $NCp7_{32-55}$ peptides (300 μM) were incubated with equimolar 2-mercaptobenzamide thioester compound in 20 mM sodium phosphate buffer, pH 7.0 at 25 °C. The UV−vis spectrum from 220 to 800 nm was collected every 0.5 h for 2.5 h on a NanoDrop spectrophotometer (NanoDrop Technologies, DE). The tetrahedrally coordinated cobalt in $NCp7_{32-55}$ has maximum absorbances at 698 and 642 nm. The absorbance at these two wavelengths was specifically monitored and compared against a control sample to which no compound was added. Each experiment was repeated a minimum of three times. Normalized absorbance values were calculated by subtracting the absorbance of the sample with compound from that of the control sample containing cobalt-refolded $NCp7_{32-55}$ without compound. The normalized absorbance at 642 and 698 nm was plotted versus time, and the initial rate of absorbance loss (0−2.5 h) was obtained from the slope of the linear regression curve.

Mass Spectrometry. Zinc- or cobalt-refolded NCp7, $NCp7_{32-55}$, or $NCp7_{32-55}$K38R was incubated with an equimolar concentration of compound 247 at 30 °C for 0−6 h. Samples were taken at various time points for further analysis. The number of total modifications was determined by MALDI-TOF MS immediately after sample removal. A second aliquot was treated with 25 mM dithiothreitol (DTT) at 4 °C for 16 h to remove labile *S*-acetyl cysteine modifications and analyzed by MALDI-TOF MS. Identification of the sites of modification was performed using the remainder of the sample. First, the modified forms

(41) Roux, C.; Gresh, N.; Perera, L.; Piquemal, J.-P.; Salmon, L. *J. Comput. Chem.* **2007**, *28*, 938−957.
(42) Srivastava, P.; Schito, M.; Fattah, R. J.; Hara, T.; Hartman, T.; Buckheit, R. W., Jr.; Turpin, J. A.; Inman, J. K.; Appella, E. *Bioorg. Med. Chem.* **2004**, *12*, 6437−6450.
(43) Basrur, V.; Song, Y.; Mazur, S. J.; Higashimoto, Y.; Turpin, J. A.; Rice, W. G.; Inman, J. K.; Appella, E. *J. Biol. Chem.* **2000**, *275*, 14890−14897.

were separated by C18 RP-HPLC with a linear gradient from 95% HPLC buffer A (0.05% TFA in water) to 65% HPLC buffer B (0.05% TFA in acetonitrile). Each peak was collected separately and analyzed by MALDI-TOF MS. For identification of cysteine modifications, fractions of interest were lyophilized and cleaved either with endoproteinase Arg-C and chymotrypsin (NCp7) or chymotrypsin and endoproteinase Glu-C ($NCp7_{32-55}$), 25 °C, 16 h. For identification of lysine modifications, fractions of interest were lyophilized, treated with 25 mM DTT at 4 °C for 16 h, and cleaved with chymotrypsin and then endoproteinase Glu-C at 25 °C for 16 h. After cleavage, the acidified peptides were separated at 500 nL/min on a 0.1 × 150 mm Magic C18AQ column (Michrom Bioresources, Inc., CA) in line after a nanotrap column using the Paradigm MS4 MDLC (Michrom Bioresources, Inc., CA). The reversed-phase separation was coupled to online analysis by tandem mass spectrometry (nLC-ESI-MS/MS) on an LTQ ion trap mass spectrometer (ThermoElectron, CA) equipped with a nanospray ion source. Elution of the peptides into the mass spectrometer was performed with a linear gradient from 95% mobile phase A (2% acetonitrile, 0.5% acetic acid, 97.5% water) to 65% mobile phase B (10% water, 0.5% acetic acid, 89.5% acetonitrile) in 20 min and then to 95% mobile phase B in 5 min. The peptides were detected in positive ion mode using a data-dependent method in which the top seven ions detected in an initial survey scan were selected for MS/MS analysis.

Circular Dichroism Spectroscopy. The fold of $NCp7_{32-55}$ wild-type and mutant peptides was assessed by measuring the circular dichroism (CD) spectrum of both unfolded and zinc-refolded $NCp7_{32-55}$ peptides (100 μM) in 10 mM sodium phosphate buffer pH 7.0. CD spectra were recorded from 190 to 260 nm on a Jasco J-715 spectrapolarimeter (Jasco, Inc., MD) at 25 °C. The spectra shown are plots of mean residue ellipticity and were corrected for background interference from the buffer.

Molecular Modeling. Modeling was carried out in two regimes: models of NCp7 and the reacting compounds were initially generated using the classical CHARMM force field, then the models were further refined by the SIBFA polarizable molecular mechanics procedure. Accordingly, for the first phase, atomic coordinates of NCp7, containing both ZD1 and ZD2, were obtained from file 1MFS.pdb of the Protein Data Bank (PDB) (http://www.pdb.org). Atomic models of compounds 247 and 295 were produced with the 2-D Sketcher and 3-D Builder applications of the Quanta-2006 molecular modeling software package (Accelrys Software Inc.). All structures were energy minimized with the CHARMM (c31b2) molecular mechanics package[44] using the "top_all22_prot" and "par_all22_prot" topology and parameter packages.[45] Investigation of possible reaction complexes was done both "manually" and with a series of automated docking simulations. The latter was conducted with the AutoDock 3.0 software package,[46] and analyzed with AutoDockTools (ADT).[47] Of these, models were judged potentially accurate if they simultaneously explained the importance of the Trp_{37} side chain (as indicated by the reaction rate mutagenesis studies) and positioned the thioester group of the compound near a sulfur atom of a coordinated sulfur atom of a ZD2 cysteine (as required for the nucleophilic attack of the proposed reaction mechanism). A representative sample of the potentially correct models was then optimized with the more quantum mechanically realistic SIBFA formalism for this type of metal-center system.[48]

1. SIBFA Polarizable Molecular Mechanics Computations. The SIBFA intermolecular interaction energy was computed as a sum of five components: electrostatic multipolar energy (E_{MTP*}), short-range repulsion (E_{rep}), polarization (E_{pol}), charge transfer (E_{ct}), and dispersion (E_{disp}). The electrostatic energy was computed using multipoles (up to quadrupoles) distributed on the atoms and the bond barycenters[49] and included penetration effects due to the overlap of the electron clouds of the interacting molecules.[50] E_{rep} was expressed as the sum of bond−bond, bond−lone pair, and lone pair−lone pair interactions and was augmented with an S^2/R^2 term.[40,50] The polarization energy was computed using anisotropic polarizabilities distributed on the centroids of the localized orbitals (heteroatom lone pairs and bond barycenters) following the procedure of Garmer and Stevens.[51] The full expression of each component has been detailed previously.[39−41] Solvation energy, ΔG_{solv}, was formulated as a sum of electrostatic, polarization, repulsion, and cavitation contributions following the Langlet−Claverie (LC) procedure.[52] The electrostatic and polarization contributions were computed with the same distributed multipoles as SIBFA, ensuring mutual consistency between the two procedures. The parameters used were the same as previously reported.[41] A high-resolution grid of 610 points per atom was used.

2. Construction of NCp7 and Compound 247. The model of ZD2 (residues 32−54 of NCp7) was assembled from its constitutive backbone and side-chain fragments in the SIBFA standard library. All three Cys residues were considered to be in the anionic form, consistent with experiment.[53] Compound 247 was assembled using the benzene, H_2S, formaldehyde, formamide, and methane standard fragments. The multipoles and polarizabilities of each fragment were taken to be those previously derived from ab initio computations using the CEP 4-31(2d) basis set[54] with standard bond lengths and valence angles. For both the protein and thioester compound, threefold torsional barriers were used to model rotations around the saturated single bonds, using the same method of calibration as previously described.[55,56]

3. Treatment of Flexible Molecules. The following is a summary of the procedures that we have previously developed.[39,41] To compute E_{MTP*}, the multipoles are redistributed along the junction bonds, X−H and Y−H, between fragments following a previously published procedure[55] to construct a connecting single bond X−Y. To compute E_{pol}, on the other hand, an alternative set of multipoles is used, for which no redistribution along the junctions is done. In this fashion, each individual fragment retains the net charge it has prior to the assembling procedure, namely, 0 if neutral, −1 if anionic, and 1 if cationic, whereas it is not retained following redistribution. This prevents an imbalance of polarization between two successive fragments that have lost their net charges, which could be amplified in the complete molecule due to the nonadditivity of E_{pol}. It was also necessary to prevent overlaps involving the H atoms belonging to the junction bonds. Such bonds were shrunk by carrying back the end H atoms on the X or Y atom whence the bond originates. Finally, upon computing the intermolecular interactions between flexible molecules, namely, ZD2 and thioester, together with their interactions with Zn(II), inter- and intramolecular interfragment interactions are computed simultaneously and consistently as a single integrated energy. This need is a consequence of the nonadditivity of E_{pol} and E_{ct}. Such an approach was recently validated upon performing parallel SIBFA and ab initio quantum-chemistry[39,41] on Zn(II)-dependent metalloenzymes.

4. Energy Minimizations. Energy minimizations were done with the Merlin package.[57] All intramolecular torsion angles of $NCp7_{32-54}$ ZD2 and compound 247 were relaxed simultaneously with the six

(44) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(45) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(46) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(47) Sanner, M. F. *J. Mol. Graphics Modell.* **1999**, *17*, 57−61.
(48) Gresh, N. *Curr. Pharm. Des.* **2006**, *12*, 2121−2158.
(49) Vigné-Maeder, F.; Claverie, P. *J. Chem. Phys.* **1988**, *88*, 4934−4948.
(50) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *J. Chem. Theory Comput.* **2007**, *3*, 824−837.
(51) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem.* **1989**, *93*, 8263−8270.
(52) Langlet, J.; Claverie, P.; Caillet, J.; Pullman, A. *J. Phys. Chem.* **1988**, *92*, 1617−1631.
(53) Bombarda, E.; Morellet, N.; Cherradi, H.; Spiess, B.; Bouaziz, S.; Grell, E.; Roques, B. P.; Mely, Y. *J. Mol. Biol.* **2001**, *310*, 659−672.
(54) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026−6033.
(55) Gresh, N.; Claverie, P.; Pullman, A. *Theor. Chim. Acta* **1984**, *66*, 1−20.
(56) Gresh, N.; Pullman, A.; Claverie, P. *Theor. Chim. Acta* **1985**, *67*, 11−32.
(57) Evangelakis, G.; Rizos, J.; Lagaris, I.; Demetropoulos, G. N. *Comput. Phys. Commun.* **1987**, *46*, 401−415.

Table 1. Rate of UV−Vis Absorbance Loss at 642 nm from $NCp7_{32-55}$ and $NCp7_{32-55}$ Mutant Peptides after Incubation with Compound 247

$NCp7_{32-55}$	rate of absorbance loss (mAU/h)
wt	15.07 ± 0.42
W37A	~0
W37F	5.67 ± 0.30
W37Y	1.73 ± 0.15
K38A	NF[a]
K38N	UNS[b]
K38R	0.48 ± 0.09
Q45K	NF[a]
Q45E	NF[a]
D48N	NF[a]
Q45ED48N	2.05 ± 0.07
M46A	13.33 ± 0.35
M46Nle	14.00 ± 0.10

[a] NF: Does not form correct tetrahedral coordination geometry. [b] UNS: Forms the correct tetrahedral coordination geometry, but metal coordination is unstable.

intermolecular variables that define the orientation of the inhibitor in its complex and the three corresponding variables for Zn(II). The first step was a round of in vacuo energy minimization with a set of distance restraints to maintain the same general complex structure as the starting model from the first phase. These included a 3.0 Å distance between the carbonyl carbon of the thioester compound and the S atom of the interacting Cys ligand, and a maximum of 5.5 Å between selected pairs of atoms of the benzoyl ring of the compound and the indole ring of Trp_{37} of $NCp7_{32-54}$. Energy minimization was then continued in vacuo without the restraints and finally in the presence of the LC continuum reaction field procedure to simulate solvent effects.

Results

An Aromatic Residue is Essential for Thioester Reactivity. As we have previously shown that the thioester reaction with NCp7 occurs only in ZD2,[38] we used mutagenesis experiments to determine if any amino acids in that domain other than those that coordinate zinc are important for the interaction and confer specificity. Only five positions differ between the two zinc-binding domains (Figure 1B): in ZD2, the residues at those positions are Trp_{37}, Lys_{38}, Gln_{45}, Met_{46}, and Asp_{48}. It is likely that one or more of these residues determine the preference of the 2-mercaptobenzamide thioester compound for ZD2 as the structure of the two domains is very similar (Supporting Information Figure 1). We focused first on Trp_{37} since we previously observed adduct formation near this residue.[38] Trp_{37} is at position $x + 1$ in the zinc-binding domain, where x is the first cysteine in the motif, either Cys_{36} in ZD2 or Cys_{15} in ZD1 (Figure 1B). Peptides were made where Trp_{37} was mutated to alanine, phenylalanine, or tyrosine (Table 1). Substitution by alanine completely changed the character of the residue. The phenylalanine mutation was introduced because phenylalanine is the residue at position $x + 1$ in ZD1 (Figure 1B), and the tyrosine mutation was made because that residue is found in position $x + 1$ in other lentiviruses, such as the bovine immunodeficiency virus.

The peptides were refolded with cobalt rather than zinc, since Co^{2+} forms a spectroscopically visible complex, whereas Zn^{2+} does not. Both metals, though, fold with the same tetrahedral geometry in the zinc-binding domain.[58] The absorbance maxima for tetrahedrally coordinated cobalt-refolded NCp7 are 642 and 698 nm.[58] Thus, by monitoring the changes in these absorbances over time, an initial rate of metal ejection can be determined and compared among the various peptides. The wild-type $NCp7_{32-55}$ peptide folded with the correct tetrahedral geometry, as shown by absorbance maxima at the expected wavelengths (Supporting Information Figure 2A), and compound 247 was able to eject coordinated cobalt with a maximal rate (Table 1). By contrast, no metal ejection was observed when cobalt-refolded $NCp7_{32-55}$W37A was incubated with compound 247, though the peptide did show tetrahedral coordination geometry (Table 1). Metal ejection was observed with both cobalt-refolded $NCp7_{32-55}$W37F and $NCp7_{32-55}$W37Y peptides at reduced rates compared with the wild-type peptide (Table 1). Cobalt was lost 3 times slower from $NCp7_{32-55}$W37F and 12 times slower from $NCp7_{32-55}$W37Y than from the native $NCp7_{32-55}$ (Table 1). This observation suggests that an aromatic residue is needed in this position for acetyl transfer to occur, though other mutations at this position do not affect the overall coordination geometry of the zinc-binding domain.

Compound 247 Reaction with NCp7 Acetylates Cys_{36} as well as Proximal Lysines. To explore sites of acetylation in ZD2 by the reaction with compound 247, mass spectrometry was used to determine where adduct formation could be observed. Full-length, synthetic, zinc-refolded NCp7 was incubated for 3 h with equimolar amounts of compound 247 and then analyzed by MALDI-TOF mass spectrometry to determine the nature of any modifications to the zinc-binding domain. This analysis revealed populations of NCp7 containing either one or two adducts of 42 Da (Supporting Information Figure 3), consistent with transfer of an acetyl group from compound 247 to NCp7. Treatment of the singly modified protein with DTT resulted in the loss of the added group (Supporting Information Figure 3). This observation is consistent with acetyl transfer from compound 247 to NCp7 at a cysteine residue because S-acetylated cysteine is labile when treated with low molecular weight thiols.[59] Treatment of the doubly modified protein with DTT resulted in the loss of only one modification (Supporting Information Figure 3), suggesting that a residue other than cysteine was also modified.

To determine the exact position of acetyl transfer to NCp7, the modified peptide was HPLC-purified to remove any unreacted protein, then digested and analyzed by tandem mass spectrometry. In the 3 h time point, cysteine adduct formation of NCp7 was found at Cys_{39}. In addition, to examine the modifications to NCp7 at earlier times, $NCp7_{32-55}$ was incubated with compound 247 for 1 h. The modified peptide was again HPLC-purified to remove unreacted protein, then digested with endoproteinase Glu-C and analyzed by tandem mass spectrometry. At the earlier time, cysteine modification was observed at Cys_{36} (Supporting Information Figure 4), though none was observed at Cys_{39}. These results suggest that the acetyl transfer between compound 247 and NCp7 is dynamic and that Cys_{36} is one of the first residues to be modified.

Interestingly, acetyl additions were also observed on five different lysine residues of NCp7 after the 3 h incubation with compound 247: Lys_{38}, Lys_{41}, and Lys_{47} in ZD2, Lys_3 in the N-terminus of the protein, and Lys_{33} in the linker (Figure 1B).

(58) Fitzgerald, D. W.; Coleman, J. E. *Biochemistry* **1991**, *30*, 5195−5201.

(59) Veit, M.; Ponimaskin, E.; Schmidt, M. F. G. Analysis of S-Acylation of Proteins. In *Posttranslational Modifications of Proteins: Tools for Functional Proteomics*; Kannicht, C., Ed.; Humana Press, Inc.: Totowa, NJ, 2002; pp 159−178.

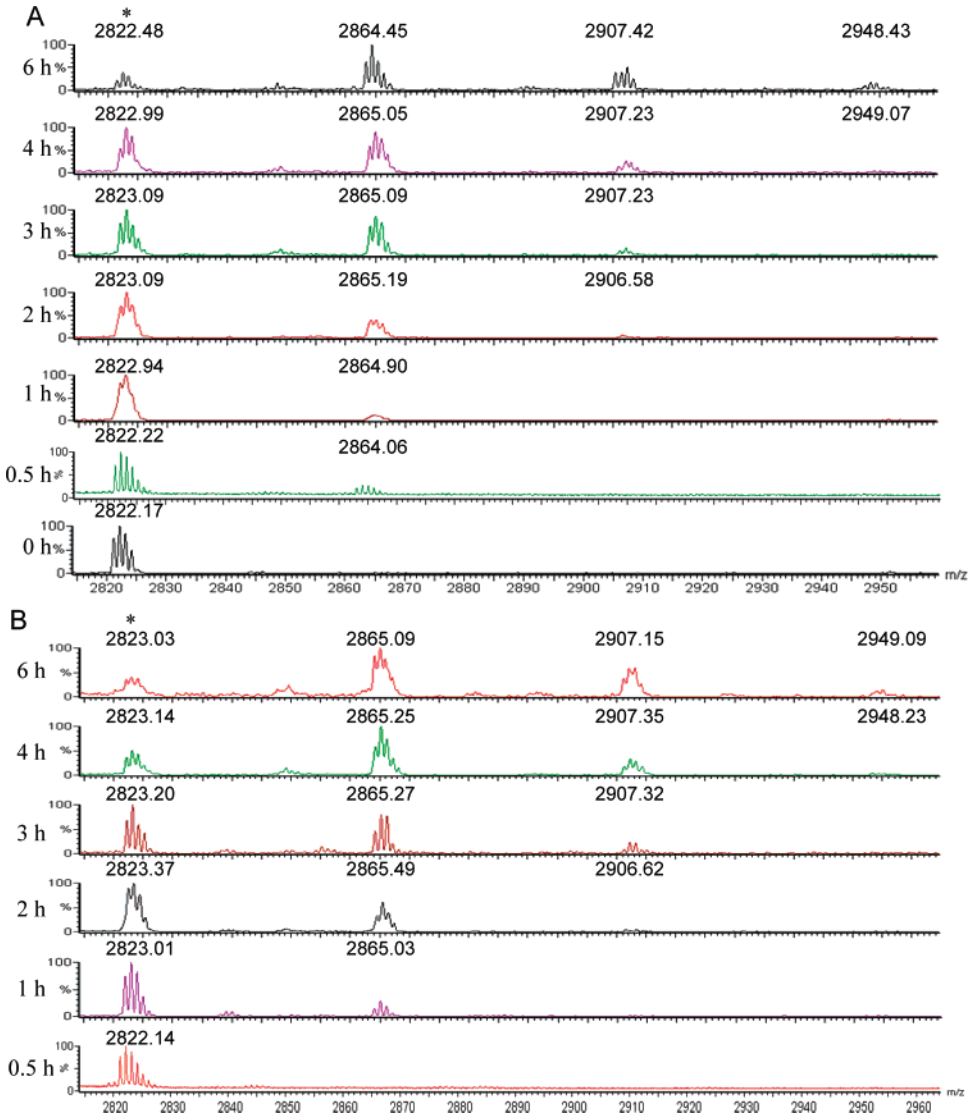


Figure 2. MALDI-TOF MS spectra collected of $NCp7_{32-55}$ incubated with compound 247 for 0−6 h (A) or after treatment with DTT (B) as described in the Experimental Methods. The mass of unmodified $NCp7_{32-55}$ is indicated with an asterisk.

No lysine acetylations were observed in ZD1 of NCp7 (data not shown). It is possible that the acetyl groups on the lysine residues form as the result of an intramolecular acetyl transfer from a modified cysteine residue (Figure 1C). While it has been shown that cysteine is the only residue able to nonenzymatically accept an acetyl group in vitro,[60] an intramolecular acetyl transfer from a cysteine to a lysine should be thermodynamically favorable and essentially irreversible.[60] Such a transfer has been previously described in 3-phosphoglyceraldehyde dehydrogenase.[61]

Intramolecular Transfer of Acetyl Groups from Cysteine Residues to Proximal Lysine Residues Occurs Rapidly after Cysteine Modification. To assess the possibility of S to N acetyl transfer, zinc-refolded $NCp7_{32-55}$ was incubated with compound 247, and samples were analyzed at various times during the incubation. Thirty minutes after thioester addition to $NCp7_{32-55}$, MALDI-TOF MS revealed the presence of a single acetyl modification to $NCp7_{32-55}$ (Figure 2A). After 2 h, both singly and doubly modified $NCp7_{32-55}$ were observed (Figure 2A), and after 3 h, a triply modified form also appeared. Over the next 3 h, there was an overall decrease in the amount of unmodified $NCp7_{32-55}$ and a concurrent increase in the modified populations, but no additional modifications were observed (Figure 2A).

Samples from each time point were further treated with DTT to remove the labile cysteine modifications. MALDI-TOF MS was then used to determine the pattern of lysine additions over time. After 1 h of $NCp7_{32-55}$ incubation with compound 247, one acetyl modification was observed (Figure 2B). After 2 h, a second acetyl addition was observable, and a third was evident after 4 h (Figure 2B). This pattern of modification is very similar to the pattern of modification without DTT treatment—in both cases, a single modification was visible after 1 h of incubation and the second modification after 2 h (Figure 2). Strikingly, though, the single modification observed after 30 min before DTT treatment was not observed afterward (Figure 2), demonstrating that the modification of NCp7 occurs first at a cysteine. After 3 h, only two lysine modifications can be detected after treatment with DTT, but three total modifications were present

(60) Dormeyer, W.; Dorr, A.; Ott, M.; Schnolzer, M. *Anal. Bioanal. Chem.* **2003**, *376*, 994−1005.
(61) Park, J. H.; Agnello, C. F.; Mathew, E. *J. Biol. Chem.* **1966**, *241*, 769−771.

before DTT treatment, further supporting the model of modification occurring first at cysteine and then transferring to lysine. When cobalt-refolded $NCp7_{32-55}$ was incubated with compound 247, a similar result was found (Supporting Information Figure 5): after 30 min, a population containing one modification was observed, and after 1.5 h a population containing a second modification was observed. After treatment with DTT, singly modified $NCp7_{32-55}$ was not observed until 1 h, and no further modification was present after 2 h (Supporting Information Figure 5). Thus, the modifications of $NCp7_{32-55}$ when refolded with either cobalt or zinc are very similar, and both patterns demonstrate that cysteine modification occurs before lysine modification.

ESI-LC MS/MS analysis was used to determine which lysine residues were acetylated. Acetyl modification on Lys_{47} was observed after 1 h and was also present after 3 and 4 h (data not shown). Transfer to both Lys_{38} and Lys_{41} was also observed after 1 h of incubation with compound 247 (Supporting Information Figure 6). Overall, lysine modification is evident early in the time course, demonstrating that the transfer from cysteine to lysine takes place rapidly after the initial reaction at a cysteine residue.

Lysine at Residue 38 Supports Optimal Reactivity. A series of mutant peptides were studied in which Lys_{38} (position $x + 2$) was mutated to alanine, asparagine, and arginine. The mutation to alanine was introduced in this position to completely change the character of the residue. The asparagine mutation changes the lysine to the residue at position $x + 2$ in ZD1 (Figure 1B), and the arginine mutation maintains a basic residue in the position. Interestingly, when $NCp7_{32-55}K38A$ was refolded with cobalt, absorbance maxima were observed at 701 and 634 nm (data not shown). The altered absorbance maxima demonstrate that replacement of Lys_{38} with alanine does not support the native tetrahedral geometry. The CD spectrum of $NCp7_{32-55}K38A$ further suggests that the peptide did not fold correctly (Supporting Information Figure 2B). $NCp7_{32-55}K38N$ was able to form a tetrahedral geometry with cobalt, with absorbance maxima observed at 698 and 642 nm, and had a CD spectrum similar to that of $NCp7_{32-55}$ (Supporting Information Figure 2). However, the coordinated cobalt precipitated out of solution after only 1 h even when no thioester compound was added. This result indicates that while $NCp7_{32-55}K38N$ is able to form the native tetrahedral geometry, the fold is not stable when coordinated by cobalt. As a result, reliable data could not be obtained for this mutant. $NCp7_{32-55}K38R$did stably coordinate cobalt, and the expected absorbance maxima at 698 and 642 nm were observed (data not shown). This demonstrates that a basic residue is required at this position for formation of a stable tetrahedral site. When $NCp7_{32-55}K38R$ was incubated with compound 247, only a very small change in absorbance was observed over 3 h. The rate of absorbance loss was approximately 30 times lower than observed with $NCp7_{32-55}$ (Table 1). The low rate of reactivity indicates that compound 247 is unable to efficiently eject coordinated metal from $NCp7_{32-55}K38R$, and thus lysine is required in this position for optimal thioester-mediated metal ejection.

Acetyl Transfer to Lysine Residues Is Important for Reactivity. The greatly reduced rate of metal ejection from $NCp7_{32-55}K38R$ could have been related to acetyl transfer with this mutant, as arginine is unable to accept the acetyl group as lysine can. To analyze the acetyl transfer reactions, compound 247 was incubated with equimolar $NCp7_{32-55}K38R$ for 6 h and aliquots were analyzed by mass spectrometry at various times without and with DTT treatment. After 1 h of incubation, a population of $NCp7_{32-55}K38R$ containing an adduct of 42 Da was observed, consistent with acetyl transfer from the thioester compound to the protein (Figure 3A). After 3 h, a small population arose in which two adducts were present. These modifications demonstrate that compound 247 was able to interact with $NCp7_{32-55}K38R$. Like the wild-type peptide, the first DTT-resistant lysine modification was observed after 1 h (Figure 3B). However, the second such modification was not observed until 4 h, later than observed with $NCp7_{32-55}$. The slower pattern of modification suggests that the lysine modification may be important in the reaction mechanism of the 2-mercaptobenzamide thioester compounds, as the loss of a single site of transfer clearly slowed the pattern of modification.

Acetyl Transfer to Lysine Residues Disrupts the NCp7 Fold. To assess the effect of the acetyl transfer on thioester reactivity, peptides were synthesized in which lysine residues of ZD2 were preacetylated. Three different peptides were analyzed based upon the observed transfer sites: $NCp7_{32-55}K38^{Ac}$, $NCp7_{32-55}K41^{Ac}$, and $NCp7_{32-55}K47^{Ac}$. Each peptide was refolded with cobalt to evaluate the rate of metal ejection caused by compound 247. Interestingly, none of the three acetylated peptides showed the expected absorbance maxima at 642 and 698 nm. The UV−vis spectrum of $NCp7_{32-55}K38^{Ac}$ had absorption maxima at 653 and 727 nm, and the spectrum of $NCp7_{32-55}K41^{Ac}$ had maxima at 655 and 700 nm and $NCp7_{32-55}K47^{Ac}$ at 687 and 717 nm (Supporting Information Figure 2A). These altered absorption maxima indicate that the acetylated peptides do not coordinate the metal with the native tetrahedral geometry. In addition, the coordinated cobalt is not stable in these peptides, and rapidly precipitates out of solution. CD spectroscopy was used to further analyze the conformation of the acetylated peptides. When refolded with zinc, wild-type $NCp7_{32-55}$ shows a small positive maximum at ∼224 nm and a decrease and red-shift in the maximum negative ellipticity from 199 to 202 nm (Figure 4A). These differences in the CD spectrum of $NCp7_{32-55}$ with and without zinc are similar to those previously described for NCp7.[6] The CD spectrum of zinc-refolded $NCp7_{32-55}K38^{Ac}$ varies significantly (Figure 4B): the positive maximum at ∼224 nm was reduced and the maximum negative ellipticity increased and showed a slight blue-shift from 202 to 201 nm. This CD spectrum demonstrates that there is a change in the conformation of $NCp7_{32-55}$ when Lys_{38} is acetylated. The CD spectra of zinc-refolded $NCp7_{32-55}K41^{Ac}$ and $NCp7_{32-55}K47^{Ac}$ show similar deviations from the wild-type spectrum (Figure 4B). Both acetylated peptides show a loss of the positive maximum at ∼224 nm and an increase in the negative ellipticity and a shift in the maximum to 201 nm. Furthermore, the CD spectrum of triply acetylated $NCp7_{32-55}$, with modifications at Lys_{38}, Lys_{41}, and Lys_{47}, was similar to that of the zinc-free peptide (Figure 4). These results indicate that transfer of the acetyl group from cysteine to any of these three lysine residues disrupts the fold of the zinc-binding domain, leading to loss of metal coordination. Therefore, it is possible that the transfer mechanism is an important part of the mechanism of NCp7 inactivation by the 2-mercaptobenzamide thioester compounds.

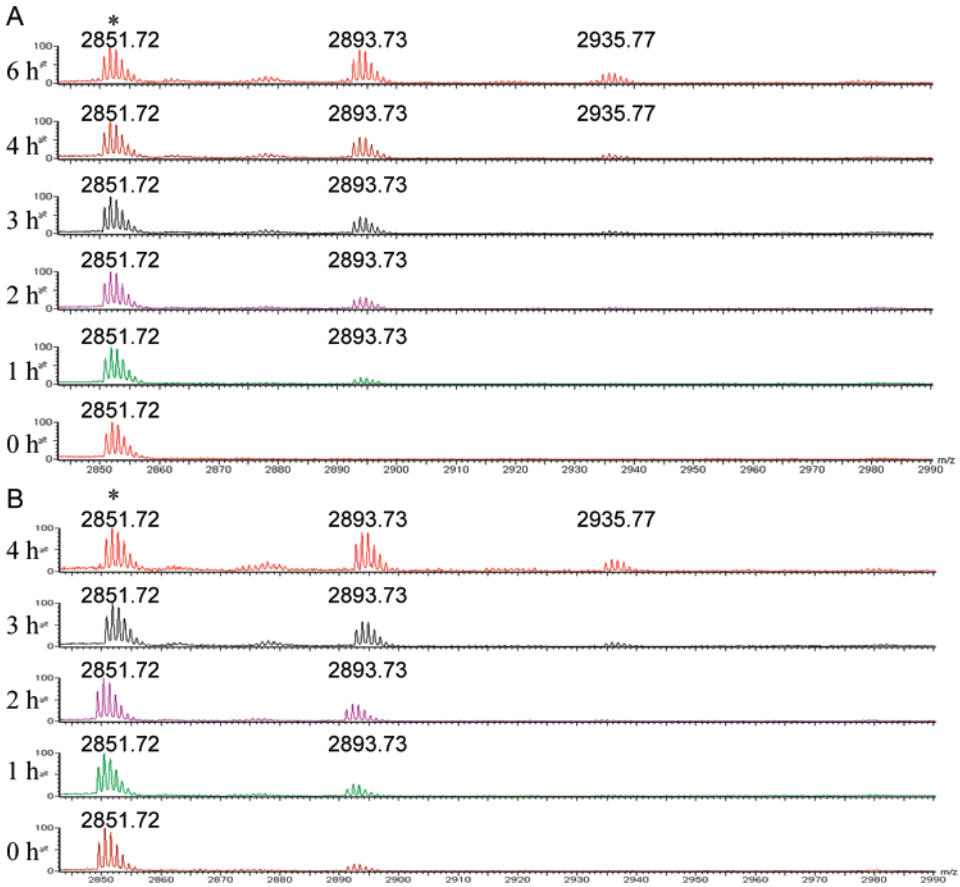


Figure 3. MALDI-TOF MS spectra of $NCp7_{32-55}K38R$ incubated with compound 247 for 0−6 h (A) or after treatment with DTT (B) as described in the Experimental Methods. The mass of unmodified $NCp7_{32-55}K38R$ is indicated with an asterisk.

Interaction with the Side Chain of Gln_{45} is Important for Full Thioester Reactivity, while the Side Chain of Met_{46} is Not. To further determine which residues are important for the specificity of the 2-mercaptobenzamide thioester compounds, the effect of altering the side chain of Gln_{45} or Asp_{48} on the rate of reactivity was studied. Gln_{45} (position $x + 9$ in ZD2, Figure 1B) is another residue for which the position varies between the two zinc-binding domains. Isoleucine, leucine, and valine are found at this position in ZD1 (Figure 5); all three residues have hydrophobic side chains that cannot make hydrogen bonds. We hypothesized that the ability of glutamine to be both hydrogen-bond donor and acceptor may play a role in the preferential reactivity of the thioester compounds. Gln_{45} was first mutated to lysine and glutamic acid to create a residue that can only act as either donor or receptor, but neither peptide formed a stable tetrahedral geometry when refolded with cobalt (Table 1). Mutation of Asp_{48} (position $x + 12$) to asparagine, the residue in the corresponding position in ZD1 (Figure 1B), also resulted in incorrect metal coordination (Table 1). The CD spectra of these mutants are also reflective of improperly folded zinc-binding domains (Supporting Information Figure 2B). The results of these mutations suggest that the balance of charges in ZD2 is very important and that any mutation that adds or subtracts a charge may result in an incorrectly folded zinc-binding domain. To circumvent this potential problem, a double-mutant peptide was designed in which Gln_{45} was mutated to Glu and Asp_{48} to Asn; this double mutant maintained the charge balance in ZD2 while altering the hydrogen-bond potential at position $x + 9$. $NCp7_{32-55}Q45ED48N$ did form a stable tetrahedral geometry when coordinated to cobalt. The reactivity of this peptide toward compound 247, though, was greatly reduced compared with $NCp7_{32-55}$ (Table 1), with metal ejection occurring 7 times slower from the double-mutant peptide than the native sequence. This loss of reactivity indicates that not only is the balance and spatial distribution of charge in the zinc-binding domain crucial for its fold but also that the specific residues at positions $x + 9$ and $x + 12$ are important for optimal reactivity.

Position $x + 10$ is another site in the motif where the amino acid sequence differs between ZD1 (Ala_{25}) and ZD2 (Met_{46}) (Figure 1B). Two Met_{46} mutant peptides were designed to assess the importance of this residue for interaction with compound 247. $NCp7_{32-55}M46A$ replaces the methionine with alanine, and $NCp7_{32-55}M46Nle$ substitutes norleucine for methionine, an isosteric modification in which CH_2 replaces the sulfur atom. Both $NCp7_{32-55}M46A$ and $NCp7_{32-55}M46Nle$ refolded with cobalt to form the expected tetrahedral geometry, yielding absorbance maxima at 642 and 698 nm. When the peptides were incubated with compound 247, metal ejection was observed at a rate similar to that of the wild-type $NCp7_{32-55}$ peptide (Table 1). Thus, replacement of methionine with either alanine or norleucine does not affect the reactivity of ZD2 with compound 247.

Docking of Compound 247 onto ZD2 Reveals an Orientation for Reaction. Having demonstrated the importance of several residues in ZD2 for optimal reaction with compound 247, we used molecular modeling to dock the compound onto ZD2. As we found that an aromatic residue is required at

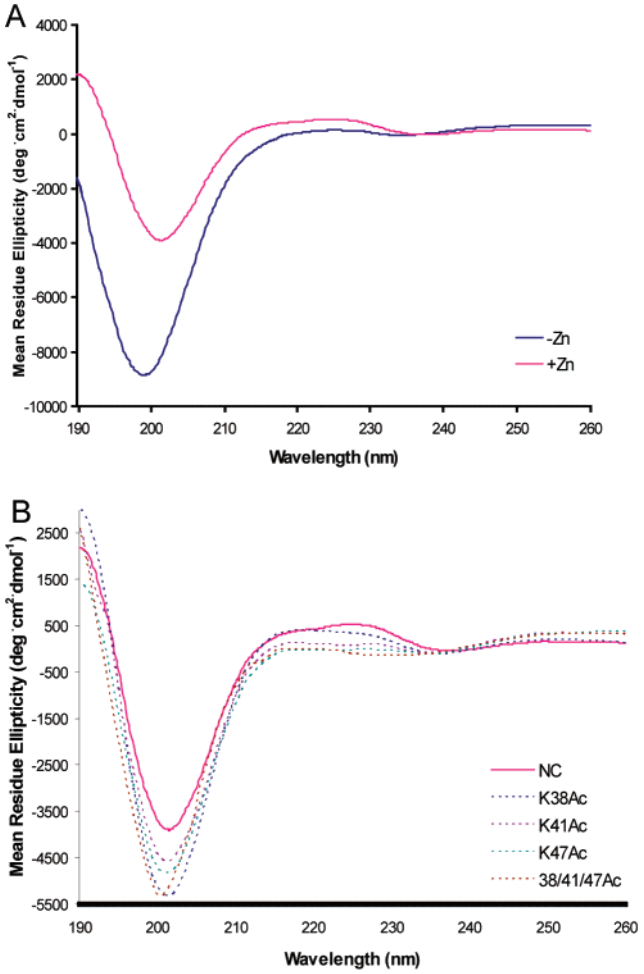


Figure 4. CD spectra of $NCp7_{32-55}$ without and with zinc (A) or zinc-refolded NCp7 mutant peptides (B).

position $x + 1$, modeling of the complex was first guided by aromatic−aromatic interactions between the indole of Trp_{37} and the phenyl ring of the thioester compound. Manual manipulation and automated docking simulations of the 3-D structures showed that this positional constraint on the surface of ZD2 limited interaction of the compound to the Cys_{36} sulfur, which lies exposed below the Trp_{37} side chain (Figure 6A). Further modeling suggested that Cys_{39} could not be the first cysteine to react since the thioester group of the compound is at least 8 Å from the sulfur of Cys_{39} while maintaining an interaction with Trp_{37} (Figure 6B). The sulfur of Cys_{39} is itself distant from Trp_{37} and buried in the protein; the C_β of Cys_{39} would have to rotate away to allow access to the sulfur.

A representative set of models resulting from the docking simulations that satisfied the criterion of simultaneous aromatic stacking and positioning of the thioester group for nucleophilic attack were selected for energy minimization using the SIBFA procedures, as described in the Experimental Methods. Of these optimized structures, the model shown in Figure 7 was among the most stable. The energy balances for compound 247 binding in this structure are provided in Supporting Information Table 1. This complex has an overall energy balance ($\delta\Delta E_{tot} + \delta\Delta G_{solv}$) of −19.2 kcal/mol, which favors its formation. It is stabilized by the second-order contributions from the polarization (E_{pol}), charge transfer (E_{ct}), and dispersion (E_{disp}) energy terms but is destabilized by the sum of first-order multipolar (E_{MTP*}) and repulsion (E_{rep}) energy contributions. The resulting in vacuo energy balance, δE_{tot}, is strongly stabilizing (−46.2 kcal/mol) but is reduced by an opposing unfavorable desolvation contribution ($\delta\Delta G_{solv}$) of 27.0 kcal/mol, resulting in a reduced but favorable energy balance. Since E_{tot} and its contributions represent actual enthalpies of binding and not free energies, the absolute magnitude of such a balance is overestimated in the absence of entropy effects that account for the reduction of translational and rotational motions of the ligand upon complex formation (for a recent discussion, see, e.g., ref 62).

Despite the release of distance constraints in the final minimization, in the most stable orientation, the phenyl ring of compound 247 maintains partial $\pi-\pi$ overlap with the Trp_{37} indole, and the carbonyl carbon of the thioester group remains sufficiently close to the sulfur of Cys_{36} for nucleophilic attack at a distance of 3.4 Å. Moreover, the angle between the carbonyl oxygen and carbon of the thioester group and the sulfur of Cys_{36} is 103.8°, in accordance with the Dunitz−Bürgi trajectory for nucleophilic reaction. Thus, the predicted geometry is compatible with the nucleophilic attack of the cysteine sulfur on the thioester. In this docking orientation, the carbonyl oxygen and amide group at the end of the β-alaninamide of compound 247 make two hydrogen bonds with the complementary donor and acceptor of the Gln_{45} side chain. The amide group of Gln_{45} makes a third hydrogen bond with the carbonyl oxygen of the compound's thioester group. This interaction is consistent with the decreased reactivity observed for $NCp7_{32-55}$Q45ED48N (Table 1). The methyl group of the acetyl portion of the compound rests in a pocket formed by the side chains of Lys_{38} and Met_{46}. Met_{46} does not interact with the acetyl group of compound 247, consistent with the full reactivity of peptides mutated at this residue (Table 1). However, manual modeling demonstrated that the side chains of these two residues can easily move away to form an extended groove, or channel, which could accommodate the binding of longer acyl groups, such as the pyridinioalkanoyl and nicotinoyl moieties present in other active 2-mercaptobenzamide thioester compounds.[38] Thus, the binding orientation presented here for compound 247 is consistent with all experimental data and could be common to all active members of this class of thioester compound.

The Side Chain of Gln_{45} Plays a Role in the Exclusion of Bulky Substituents to the Thioester Compound. As described above, mutational data demonstrated the importance of Gln_{45} and the docking model suggests that this residue may play a role in selection of reactive thioester groups. We hypothesized that replacement of Gln_{45} with α-aminobutyric acid (Abu) would allow greater variation in the substituents on the 2-mercaptobenzamide thioester compound as Abu contains a small, nonpolar side chain. $NCp7_{32-55}$Q45Abu formed a stable, tetrahedral fold when coordinated with cobalt, but when incubated with compound 247, the rate of metal ejection was 3 times lower than that of the wild-type peptide (Table 2), possibly related to a loss of stabilizing hydrogen bonds. The reactivity of $NCp7_{32-55}$Q45Abu with compound 295 (Figure 1A) was also measured; in this compound, the amide group at the end of the β-alaninamide was replaced with a methyl ester group. Closely related compounds, terminating in an ester rather than an amide group, have previously been shown to have no antiviral activity.[37] As expected, when compound 295 was incubated with

(62) Wu, H. J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825−6830.

```
                                        ZD1                 ZD2
                  Consensus MQrgnfkg kr vKCFNCGkeGHiArNCrAPRKkGCWKCGkEGHQMKdCterQAN
A.KE.98.M264_gag_AY772970 --KGNFRGQKR I------KE--L-R--R----K------K------D-TER---
A1.TZ.97.97TZ03_AF361873 --RGNFRGQKR I------KD--L-R--R----K------K------D-TER---
B.AU.93.MBC18_AF042102 --RSNFRNQKRNV------KE--I-R--R----R------K------D-TER---
B.CN.02.02HNsq4_AY275555 --RSNFRSQRKTV------RE--I-K--K----K------K------D-TER---
C.BR.04.04BR013_AY727522 --RGNFKGPKRPI------KE--L-K--R----K------K------D-TER---
C.ZA.98.TV002_AY162224 --RSNFKGPRRAV------RE--I-R--R----K------K------D-TER---
D.KE.98.M389_16_AY772980 --RGNFNGPKKII------KV--I-K--R----R------R------E-TEA---
O.FR.92.VAU_AF407418 --RGNKPIRKGPI------KE--V-K--R----K------Q------D-KNR---
```

Figure 5. Consensus sequence of NCp7 across the strains of HIV-1. The sequence data is from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov). Below the consensus sequence are shown select sequences as examples of the sequence variation. In the consensus sequence, residues with >95% identity are shown as capital letters and those with 51−95% identity are shown in lowercase. The residues that make up the two zinc-binding domains are marked, and in red are the residues we show to be important for the thioester reactivity.

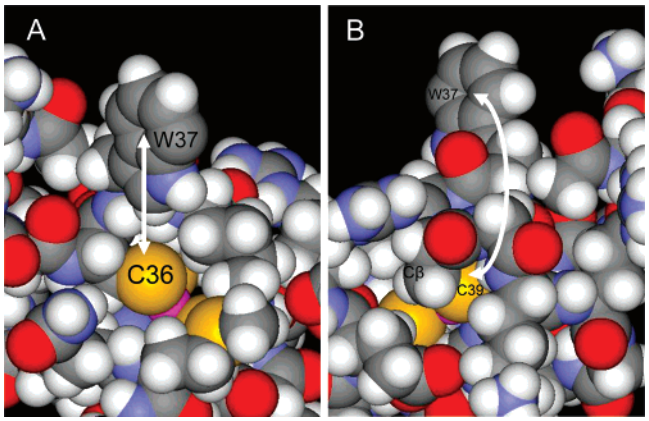


Figure 6. Space-filling model of the C-terminal ZBD of NCp7 (ZD2) showing access to Cys_{36} (A) and Cys_{39} (B). The color code is the following: carbon, gray; oxygen, red; nitrogen, blue; hydrogen, white; sulfur, yellow; zinc, magenta. The side chain of Trp_{37}, the sulfur atoms of Cys_{36} and Cys_{39}, and the β-methylene of Cys_{39} are labeled. The shortest path between the indole side chain and the cysteine sulfur in each panel is illustrated by a white arrow.

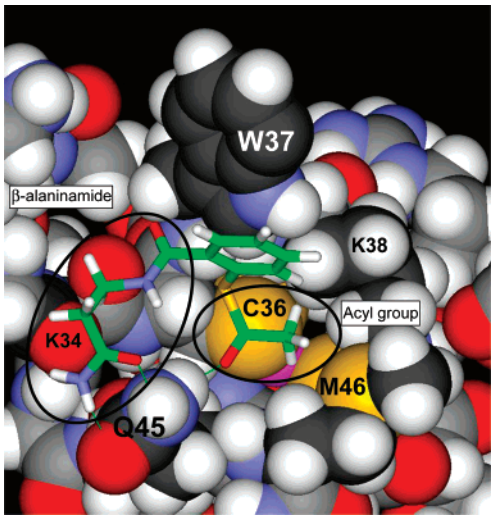


Figure 7. Representative model of the complex of ZD2 (space-filling) and compound 247 (sticks). Except for the following exceptions, the color code is the same as in Figure 6. For distinction, the carbon atoms of the compound are green. In addition, ZD2 peptide residues at the interface and mentioned in the text are highlighted by a darker shade of gray for the carbons as well as by labels. Intermolecular hydrogen bonds are shown as green lines. The acyl and β-alaninamide groups of the compound are also indicated.

cobalt-refolded $NCp7_{32-55}$, no metal ejection was observed (Table 2). However, when compound 295 was incubated with $NCp7_{32-55}$Q45Abu, metal ejection was observed at the same rate as observed for compound 247 (Table 2). Thus, interactions with Gln_{45} are important for reactivity and help determine which 2-mercaptobenzamide thioester compounds are able to react with ZD2.

Table 2. Rate of UV−Vis Absorbance Loss at 642 nm from $NCp7_{32-55}$ and $NCp7_{32-55}$Q45Abu Peptides

peptide	247	295
$NCp7_{32-55}$	15.07 ± 0.42	0
$NCp7_{32-55}$Q45Abu	4.53 ± 0.18	3.20 ± 0.53

Discussion

We have previously reported on the mechanism of action of the 2-mercaptobenzamide thioester compounds with zinc-binding domains.[38] We proposed that the thiol of a zinc-coordinated cysteine attacks the carbonyl carbon of the thioester, resulting in an acyl transfer from the thioester to the cysteine sulfur, and loss of zinc coordination (Figure 1C).[38] We demonstrated that reaction with the thioester compounds results in the inability of NCp7 to bind nucleic acid.[38] We have also observed that the rate of reaction varies widely among zinc-binding domains that have the same general motif.[38,63] NCp7 is a good example of the diversity observed among zinc-binding domains—the protein contains two domains with the same coordination motif and spacing (Supporting Information Figure 1), yet one domain is reactive while the other is not.[38,63] Clearly, the CCHC coordination motif per se (Figure 1B) cannot account for trends in zinc-binding domain reactivity. However, analysis of experimental NCp7 protein structures reveals that the local protein environments of the zinc-coordinated cysteines, contained within ZD1 and ZD2, are not equivalent, where the cysteines of ZD2 are significantly less shielded or stabilized by the surrounding protein than ZD1,[64] consistent with observations of NCp7 reactivity.[31,38,65] Moreover, the amino acids surrounding the zinc-coordinating residues of ZD2 may confer specific NCp7 interactions with the 2-mercaptobenzamide thioester compounds, resulting in site-directed inhibition of ZD2. For example, in the alkylation of the zinc-binding domain of

(63) Jenkins, L. M. M.; Durell, S. R.; Maynard, A. T.; Stahl, S. J.; Inman, J. K.; Appella, E.; Legault, P.; Omichinski, J. G. *J. Am. Chem. Soc.* **2006**, *128*, 11964−11976.
(64) Maynard, A. T.; Covell, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 1047−1058.
(65) Chertova, E. N.; Kane, B. P.; McGrath, C.; Johnson, D. G.; Sowder, R. C., II; Arthur, L. O.; Henderson, L. E. *Biochemistry* **1998**, *37*, 17890−17897.

the Ada DNA repair protein, structural studies have revealed that the absence of an intramolecular hydrogen bond to the reactive cysteine accounts for its enhanced nucleophilicity compared with the other cysteine residues in the domain.[66,67] Thus, we used mutagenesis experiments to identify residues in ZD2 that are important for reaction with the 2-mercaptobenzamide thioester compounds.

We determined that position $x + 1$ (where x is Cys_{36}) needs to be an aromatic residue for reactivity, with tryptophan preferred over phenylalanine, and the latter preferred over tyrosine (Table 1). While a basic residue (lysine or arginine) is required at position $x + 2$ for the correct fold, a lysine residue is needed for full reactivity (Table 1). Also, we have shown that the presence of a hydrogen-bond donor in position $x + 9$ (glutamine) is important for optimal reactivity of the thioester compound (Tables 1 and 2). Thus, we have identified three key residues that modulate thioester reactivity in ZD2, which provide geometrical constraints for uniquely defining the putative docking orientation of compound 247. Interestingly, these three residues in particular, and many residues in ZD2 in general, are highly conserved among HIV-1 strains (Figure 5). This indicates that these residues are important in NCp7 structure and function. Thus, the 2-mercaptobenzamide thioester compounds capitalize on critical residues that are unlikely to be mutated.

The results presented here have enabled a better understanding of the reaction mechanism of these thioester compounds (Figure 1C). The 2-mercaptobenzamide thioester compound likely first docks onto NCp7 via an aromatic, $\pi-\pi$ interaction between the indole side chain of Trp_{37} and the phenyl ring of the thioester compound. This interaction, combined with hydrogen bonding of the β-alaninamide to the Gln_{45} side chain, optimally orients the thioester for nucleophilic attack by the sulfur atom of Cys_{36}. An acyl transfer then takes place from the thioester to the cysteine sulfur, resulting in a new thioester linkage on the protein and release of a thiol group (Figure 1C). Here, we have demonstrated that a second, intramolecular acyl transfer occurs from the cysteine to the Nϵ of a proximal lysine residue (Figure 1C). The intramolecular S to N transfer is irreversible, since the acyl group is thermodynamically more stable on the lysine. The transfer of the acyl group neutralizes the charge on the lysine residue and leads to unfolding of ZD2. This may then initiate unfolding of ZD1, resulting in the total loss of NCp7 structure and function.[38,43]

Because of the instability of the modification on the cysteine residues, the transfer mechanism is likely to be a crucial part of the reaction mechanism of the thioester compounds. Indeed, we observed that when Lys_{38} was mutated to arginine, which cannot act as an acyl acceptor, the ability of compound 247 to eject coordinated metal was greatly reduced (Table 1). Further, a slower pattern of modification was observed with this mutant (Figure 3B). It is likely that the slower pattern of modification is caused by the inability of the acetyl modification to transfer to residue 38 when it is arginine. Four other lysine residues remain in this mutant though, providing alternative sites for transfer, as evidenced by the DTT-resistant modified populations. The slower pattern of modification observed by mass spectrometry, combined with the small overall amount of modified protein (<1%, data not shown), could have resulted in the greatly reduced rate of metal ejection from cobalt-refolded $NCp7_{32-55}$K38R. Mass spectrometry is more sensitive than the UV−vis spectroscopy and able to detect a small amount of modified protein which may not have as large an impact on the absorbance. The observed reduction in reactivity demonstrates the importance of the S to N transfer in the reaction mechanism. Many of the lysine residues where acyl transfer was observed have been shown to contact RNA in NCp7:RNA complexes.[15,16] This indicates that acyl transfer to these residues would impair or block RNA binding, preventing NCp7 from functioning. Furthermore, acyl transfer to Lys_{38}, Lys_{41}, or Lys_{47} alters the tetrahedral geometry of NCp7 (Table 1 and Figure 4), and acetylation of all three residues results in a conformation similar to that of the metal-free protein (Figure 4). Together, these results indicate that acyl transfer to these residues in the course of the reaction with the 2-mercaptobenzamide thioesters would disrupt the fold of the protein. In addition, these three lysine residues are highly conserved in NCp7 sequences across viral strains, emphasizing their importance in the protein (Figure 5). The stability of the acylation on the lysine residue suggests that this second modification is the one that inactivates NCp7, while cysteine acylation gives a more transient intermediate.

The lack of a reaction at ZD1 can be understood, in part, by the preference of the thioester compound for the indole side chain of ZD2 at the $x + 1$ position. That is, when Trp_{37} was replaced by Phe, the residue in the corresponding position of ZD1, the reactivity of compound 247 was 3 times less (Table 1). Interestingly, the aromatic−aromatic interactions we propose for the initial binding of 247 to NCp7 have been similarly observed for NCp7 binding to RNA. The indole of Trp_{37} stacks with the aromatic base of a guanine nucleotide when NCp7 binds to ΨSL2 or ΨSL3 to package the viral genomic RNA.[15,16] The inability of NCp7 complexed with RNA to interact with the thioester compounds[38] may be related to the decreased accessibility of Trp_{37} in the complex, as well as increased steric occlusion of coordinated cysteines within ZD2. Likely because of its important functional role, Trp_{37} is found to be nearly invariant among different strains of HIV-1 (Figure 5). Thus, it is unlikely that a drug-resistant mutation at residue 37 would arise in viral strains exposed to the 2-mercaptobenzamide thioester compounds.

The mutational studies presented here also demonstrate the importance of residues relatively distant from the zinc coordination site ($\geq$7 Å) to the binding specificity. In particular, the steric and hydrogen-bonding potential of the side chain at residue 45 in ZD2 must complement the functional group at the end of the thioester. We observed that changing Gln_{45} to Glu significantly reduced the rate of reaction (Table 1), possibly due to the loss of a hydrogen-bond donor to the terminal amide group of the thioester compound. Similarly, the observed lack of activity of compound 295 with native NCp7 (Table 2) may be due to the inability of the methyl ester group to accept hydrogen bonds. Additionally, the bulkiness of the additional methyl group may sterically prevent a favorable hydrogen bond from forming. That the Q45Abu mutation partially restored activity for compound 295 indicates that both steric and hydrogen-bonding factors are important. Finally, the importance of a hydrogen bond at this site is further supported by the previous observation

(66) He, C.; Hus, J. C.; Sun, L. J.; Zhou, P.; Norman, D. P. G.; Dotsch, V.; Gross, J. D.; Lane, W. S.; Wagner, G.; Verdine, G. L. *Mol. Cell* **2005**, *20*, 117−129.
(67) Penner-Hahn, J. *Curr. Opin. Chem. Biol.* **2007**, *11*, 166−171.

that thioester compounds retained activity when the amide group at the end of the 2-mercaptobenzamide group was monosubstituted but not when it was disubstituted.[42]

Previously, we identified Cys_{39} as the site of reaction with the thioester compounds.[38] This interpretation was based on mass spectrometry results collected after 3 h, which identified only this cysteine residue as modified by acylation, a result that is reproduced in this study. When analysis was performed on $NCp7_{32-55}$ after 1 h of incubation, acetylation was observed on Cys_{36} and not Cys_{39}. The current analysis of the 1 h time point required usage of the single-finger domain, rather than the double-finger protein, to generate enough modified peptide to study. The lack of observable acetylation at Cys_{36} in the 3 h sample may be caused by the relatively fast transfer of the acetyl group to a proximal lysine, occurring long before being detected by mass spectrometry. It is also possible that peptides in which Cys_{36} is acetylated are present at that time, but not as readily observed by mass spectrometry as acetylated Cys_{39} peptides. Our previous NMR analysis indicated a similar rate of signal intensity decrease for Cys_{36} as Cys_{39},[38] but this experiment does not provide information on the site of acetylation. The results of this study suggest that Cys_{36} acetylation occurs before Cys_{39} acetylation, though both modifications occur early in the reaction time scale. Thus, the model developed based upon the mutagenesis studies, in which Cys_{36} was in a closer position for reaction, is consistent with the results of the shorter incubation.

Protein inhibition based upon covalent modification is not a new mechanism for drugs. Rather, the efficacy of several common drugs is based upon their ability to covalently modify their protein target. Penicillins irreversibly acylate transpeptidase or carboxypeptidase, forming an inactive complex and blocking bacterial cell wall biosynthesis.[68] Aspirin functions via the selective acetylation of a serine residue near the carboxyl terminus of cyclooxygenase.[69] Omeprazole, a proton pump inhibitor, blocks the activity of gastric ($H^+ + K^+$) ATPase by reaction with essential thiols on the protein and formation of intermolecular disulfide bonds.[70] Furthermore, the inhibitory activity of omeprazole is due to its access to the protein, not because of a specific binding site on the protein.[71] This mode of inhibition is similar to that of the 2-mercaptobenzamide thioester compounds, in which the access to the reaction site is more crucial than a high binding constant. Thus, potent, safe inhibitors do not have to bind tightly to the protein but can act via highly selective covalent modification of the protein target.

We have shown that the 2-mercaptobenzamide thioester compounds do have amino acid sequence selectivity, requiring aromatic and lysine residues proximal to the reactive cysteine, as well as positional stabilization at multiple points on the molecule. Indeed, when these three residues were mutated into ZD1, its reactivity doubled (L.M.M.J., unpublished data). Also, we have demonstrated that the reaction mechanism of the thioester compounds consists of two separate acyl transfers, one intermolecular transfer from the compound to the cysteine sulfur and a second intramolecular transfer from the cysteine to a proximal lysine. The second transfer is the more stable one, and likely represents the final mechanism of inactivation of NCp7. These compounds are very different from previous compounds designed to target NCp7, which act by thiol−disulfide interchange initially and primarily interacting with Cys_{49}.[28,65] Studies on the pK_a of the zinc-coordinating cysteines have demonstrated that Cys_{49} has the highest pK_a value and coordinates zinc most weakly,[53] and so, it is possible that these compounds react with the most available cysteine in the protein, consistent with modeling studies.[64] By contrast, we have shown that the 2-mercaptobenzamide thioester compounds require specific and highly conserved residues in proximity to cysteine for reactivity, demonstrating the specificity of these compounds. The specificity of these interactions suggests the design of site-directed inhibitors that act via covalent modification of ZD2. Key interaction features define a structural template for pursuing a new generation of HIV-1 inhibitors, targeting a critical viral protein (NCp7) that is resistant to mutation.

Acknowledgment. The authors thank Drs. Daniel Appella and Andrew Maynard for critical review of the manuscript and helpful discussion. This research was supported in part by the Intramural Research Program of the NIH, NCI, and NIAID. The SIBFA computations were performed on the computers of the Centre de Calcul Recherche et d'Enseignement (CCRE) of the Universite Pierre-et-Marie-Curie (Paris, France), the Centre de Ressources Informatiques de Haute Normandie (CRIHAN, France), and the Centre d'Informatique National de l'Enseignement Superieur (CINES, Montpellier, France).

Supporting Information Available: Complete refs 23, 26, and 45, chemical identification of compound 247, Supporting Figures 1−6, Supporting Table 1, and the PDB coordinates for the model shown in Figure 7. This material is available free of charge via the Internet at http://pubs.acs.org.

JA071254O

(68) Waxman, D. J.; Strominger, J. L. *Annu. Rev. Biochem.* **1983**, *52*, 825−869.
(69) Vane, J. R.; Botting, R. M. *Thromb. Res.* **2003**, *110*, 255−258.
(70) Lorentzon, P.; Eklundh, B.; Brandstrom, A.; Wallmark, B. *Biochim. Biophys. Acta* **1985**, *817*, 25−32.
(71) Keeling, D. J.; Fallowfield, C.; Underwood, A. H. *Biochem. Pharmacol.* **1987**, *36*, 339−344.