Probing the Topography of HIV-1 Nucleocapsid Protein with the Alkylating Agent *N*-Ethylmaleimide†

Elena N. Chertova,* Bradley P. Kane, Connor McGrath, Donald G. Johnson, Raymond C. Sowder, II, Larry O. Arthur, and Louis E. Henderson

AIDS Vaccine Program, SAIC Frederick, National Cancer Institute, Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702

Received April 21, 1998; Revised Manuscript Received September 28, 1998

ABSTRACT: Retroviral nucleocapsid (NC) proteins contain one or two zinc fingers (ZFs) consisting of a CCHC peptide motif that coordinates Zn(II). Mutational and biochemical analyses have shown that NC ZFs are directly involved in multiple stages of viral replication, including genomic RNA encapsidation, virus maturation, and the early infection process. The multiple roles of the conserved retroviral ZFs make them attractive targets for antiviral agents. We have previously shown that a variety of chemical compounds can inactivate the whole virus by attacking NC ZFs. For the enhancement of the specificity of antiviral reagents, it is desirable to have a detailed knowledge of the spatial organization of reactive sites on the NC protein in its free and oligonucleotide-bound states. A method has been developed using chemical probes to assess the reactivity of specific Cys residues in the NC protein, and is being used to investigate the topography of ZFs in different contexts. In this study we focus on the reaction mechanism of *N*-ethylmaleimide (NEM) with free HIV-1 NCp7 protein. Our results show that the conformation of free NCp7 restricts the initial site of attack to Cys-49 (the most distal Cys residue in the second ZF) and that the reactivity of thiols in full-length protein differs from that of the isolated ZF peptides. A moderate to near complete reduction in reaction rate was observed when NCp7 was complexed with different oligonucleotides. These findings provide a set of experimentally determined parameters that can serve to guide computational modeling of the NC protein and will be useful for the rational design of drugs directed against retroviral ZFs.

Retroviral nucleocapsid (NC[1]) proteins contain one or two zinc fingers (ZFs) consisting of a CCHC peptide motif that coordinates Zn(II) through His imidazole and Cys thiolate groups (*1*, *2*). The Gag precursor of HIV-1 contains two ZF structures located in the NC domain. Gag precursors with competent zinc-binding fingers are not required for assembly and budding of a virion-like particle but are required for recognition and packaging of genomic RNA into the particle (*3*−*9*). After maturation proteolysis, both Gag ZFs are located in the mature NCp7 protein where they interact with packaged RNA in the core of the virus. ZFs in the mature NCp7 protein also are required in the infectious process (*7*, *9*). Some mutants with altered ZFs have been shown to package RNA but are severely impaired during the infectious process (*10*). These observations clearly show that the NC protein is required during the infectious process, but its exact biological role is still under investigation. The mature NC protein has been found to accompany the RNA in the preintegration complex (*11*), where it may function to facilitate nucleic acid rearrangements associated with reverse transcription and possible other steps leading to integration. Although all of the biological functions of ZFs in the NC protein are not yet clearly understood, there is little doubt that they play a vital role in the management of nucleic acids during the assembly, maturation, and infectious stages of the viral replication cycle.

ZFs in the NC protein are susceptible to chemical attack by electrophilic reagents, including disulfides and maleimides (*12*−*14*). These reagents react with cysteine thiols in the zinc coordination complex, causing chemical modifications that alter the metal-binding sites with subsequent loss of the biologically active peptide conformation. These observations suggested that the chemical reactivity of the NC ZFs might form the basis for a new class of antiviral drugs. We and others have identified several reagents (mostly disulfides) that can attack thiols in the NC protein. However, it is not known if any possess a specificity for the viral target. Nevertheless, it might be possible to find or design reagents with specificity for the viral target. As a first step in the design process, we have elected to study the detailed reaction pathway for a few selected compounds reacting with the ZFs

† This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000. Research sponsored by the National Cancer Institute, Department of Health and Human Services (DHHS). The contents of this publication do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. Government.

* To whom correspondence should be addressed. Elena N. Chertova, Ph.D., Protein Chemistry Laboratory, AIDS Vaccine Program, SAIC Frederick NCI-FCRDC, P.O. Box B, Building 535, Room 424, Frederick, MD 21702. Phone: (301)-846-1455. Fax: (301)-846-7119. E-mail: chertova@avpvx1.ncifcrf.gov.

[1] Abbreviations: NC, nucleocapsid; NEM, *N*-ethylmaleimide; NCp7, nucleocapsid protein 7; ZF, zinc finger; MS, mass spectrometry; H_2O, water; Cys-NEM, *S*-alkylated derivative; TFA, trifluoroacetic acid.

10.1021/bi980907y CCC: $15.00
Published on Web 12/01/1998

in the HIV-1 NCp7 protein and in protein−nucleic acid complexes.

Here we report on reactions of *N*-ethylmaleimide (NEM) with HIV-1 NCp7. The reagent is well-known for its reactivity with thiols in general and reacts with cysteine to form a stable, modified derivative (Cys-NEM) that can be identified by standard amino acid analysis and Edman degradation. A method has been developed using limited amounts of NEM to assess the reactivity of specific Cys residues in NCp7 protein, and is being used to investigate the topography of free protein and protein bound to oligonucleotides. The chemical stability of NEM-modified cysteine is a major factor in the analysis of the reaction products and allows us to identify the most reactive cysteine thiols in the ZFs of HIV-1 NCp7. The various NEM-modified protein intermediates can be separated by reversed-phase HPLC and identified by protein chemistry. Once standardized, the HPLC elution profile can be used to determine the amounts of unreacted protein and various reaction intermediates in a given sample. The reaction conditions can be controlled by limiting the amount of NEM and by varying reaction times such that initial reaction products are favored.

MATERIALS AND METHODS

Oligonucleotides were synthesized by Marilyn Powers, NCI-FCRDC, Frederick, MD. *N*-Ethylmaleimide purchased from Aldrich was dissolved as 100 and 1 mmol stocks in UV grade H_2O (Water; EM Science) and was stored at −20 °C. β-Mercaptoethanol was from Sigma. Acetonitrile and H_2O (UV grade) were obtained from EM Science; trifluoroacetic acid (TFA, HPLC/Spectra grade) was from Perkin-Elmer (Applied Biosystems Division, Warrington, Great Britain). Tris (ULTROL Grade) was from Calbiochem, La Jolla, CA.

HIV Recombinant-NCp7. HIV-1 NC protein was cloned into the pMal-c vector, expressed as a fusion protein in *Escherichia coli* bacteria and purified at the AIDS Vaccine Program, NCI-FCRDC, Frederick, MD (*15*). The final purification was carried out by reversed-phase HPLC on μBondapak C_{18} 25 × 100 mm column (Millipore, Inc.) at flow rate of 10 mL/min. After reduction of the protein with 2% β-mercaptoethanol for 30 min at 20 °C with stirring under nitrogen, TFA was added to reduce the pH to 2.0. The gradient was 10−24%, 130 min; 24−60%, 150 min; and 60%, 10 min. Peaks were detected by a LKB 2140 rapid spectral recoder at 206 nm to detect the absorbance of peptide bonds, 280 nm to detect the presence of aromatic amino acid residues, and 260 nm to detect nucleic acid. Two equivalents of Zn^{2+} were added per nucleocapsid protein, and the sample was lyophilized. The protein was analyzed by amino acid sequencing on an automated Applied Biosystems Inc. 477 A protein sequencer, and by amino acid analysis on an automated Beckman System 6300 analyzer. To confirm that the ZF structure had spontaneously formed, we dissolved the lyophilized sample in a pH 7.0 buffer (20 mM Tris-HCl) and analyzed it by NMR at the University of Maryland, UMBC, MD.

Reaction of Free and Oligonucleotide-Bound NCp7 with NEM. HIV-1 r-NCp7 protein, 7.5 μM, either containing 15 μM $ZnCl_2$ or free of metal ions, was reacted with a 6-fold excess of *N*-ethylmaleimide at pH 7.0 (20 mM Tris-HCl) at 37 °C. To obtain the oligonucleotide-bound state, we incubated Zn^{2+}-NCp7 with different four or eight desoxyribonucleotides at a protein−nucleotide ratio of 1:4 or 1:2 at pH 7.0 for 5 min prior to the addition of NEM. The reaction products were separated by reversed-phase HPLC on a Vydac $5C_{18}$-300 column (2.1 × 150 mm) at 0.3 mL/min using a Shimadzu HPLC system equipped with LC-10AD pumps, SCL-10A system controller, CTO-10AC oven, FRC-10A fraction collector, and SPD-M10AV diode array detector. The gradient of buffer B (0.1% TFA in acetonitrile) was the following: 0−16%, 5 min; 16−25%, 35 min; 25−80%, 10 min; 80%, 5 min. Peaks were detected at 206, 260, and 280 nm. After separation, fractions were characterized by sequencing using an automated Applied Biosystems Inc. 477 A protein sequencer.

Proteolytic Digestion. The samples of either NCp7 or modified protein were dissolved in 0.02 M Tris-HCl, pH 7, and Arg-C endoproteinase (Boehringer Mannheim Gmbh, Mannheim, Germany) (50:1 w/w NC protein−protease) was added and incubated for 1 h at 37 °C. Then, chymotrypsin (TLCK-treated) (Sigma, St. Louis, MO) was added at the same ratio as Arg-C and incubation was continued for 2 h at room temperature. After digestion, all samples were analyzed by reversed-phase HPLC on an α-Chrom $5C_{18}$-300 (2.0 × 100 mm) column at 0.3 mL/min. The gradient of buffer B was the following: 0−16%, 5 min; 16−25%, 20 min; 25−80%, 10 min; 80%, 5 min. Peaks were detected by UV absorption at 206 and 280 nm and analyzed by sequencing using an automated Applied Biosystems Inc. 477 A protein sequencer.

Fluorescence Study. All fluorescence measurements were performed on a SPEX Fluoromax spectrofluorimeter, with excitation set at 280 nm and emission at 356 nm (1 and 8 nm slit width, respectively). Measurements were performed at 37 °C in a 2 × 10 mm path length quartz cuvette (Uvonic Inc.). p7 coordinated with 2× Zn^{2+} was dissolved in 20 mM Tris-HCl buffer (pH 7.0) at 2 μM concentration. NCp7 initial fluorescence was monitored for 2.5 min after which NEM at a protein−NEM ratio 1:200 was added to the cuvette and its decrease in fluorescence was monitored over time.

MALDI-TOF MS. Modified protein was analyzed by MALDI-TOF MS using a Shimadzu Kompact Maldi-II laser desorption mass spectrometer. RP-HPLC purified fractions were lyophilized and then dissolved in 3 μL of water. The samples were prepared for MALDI-TOF MS analysis by mixing 1 μL of sample protein with 0.3 μL of matrix (sinapinic acid, 10 mg/mL, in 50% water, 50% acetonotrile) directly on the stainless steel target. Laser desorption−ionization mass spectra were acquired with low laser fluence; each spectrum was the sum of at least 100 laser shots. Substance P (M_r 1349.6 kDa), bovine insulin (M_r 5734.6 kDa), and ubiquitin (M_r 8565.9 kDa) were used to calibrate the MALDI-TOF MS instrument.

RESULTS

Modification of HIV-1 NCp7 by NEM. NEM is an alkylating agent that reacts with cysteine thiols to generate a stable S-alkylated derivative (Cys-NEM) which can be identified by Edman degradation. Preliminary studies showed that NEM reacts with thiols in NCp7. Since NCp7 contains six Cys residues (three in each ZF), it was of interest to

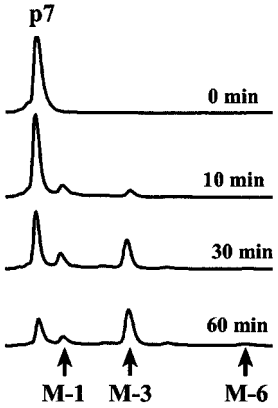

FIGURE 1: Time study of *N*-ethylmaleimide modification of HIV-1 NCp7. The protein, 2 μg per analysis, at 7.5 μM was incubated with 6-fold excess of NEM in 0.02 M Tris HCl, pH 7.0, at 37 °C. After the reaction for 0, 10, 30, and 60 min, samples were separated by HPLC on a Vydac $5C_{18}$-300 column (2.1 × 150 mm) at 0.3 mL/min. The gradient of buffer B was the following: 0−16%, 5 min; 16−25%, 20 min; 25−80%, 10 min; 80%, 5 min. Peaks were detected at 206, 260, and 280 nm. P7 is unreacted protein. M-1, M-3, and M-3 are proteins with one, three, and six NEM-modified Cys residues, which was confirmed by MALDI-TOF MS.

Table 1: Matrix-Assisted Laser Desorption-Ionization Mass Spectrometry (MALDI-MS) of p7 and Modified with NEM Protein

	measured mass (Da)	calculated mass (Da)
P7	6458.96	6451.5
M-1	6570.5	6576.63
M-3	6814.1	6826.89
M-6	7210.6	7202.28

determine if the thiols in the native protein reacted in any specific order. For this study we used rate-limiting concentrations of NEM to hold the extent of the reaction to its initial products and to slow the rate of the reaction so that initial products might be identified. We selected a protein concentration (7.5 μM) to be compatible with subsequent analysis and added NEM in 6-fold molar excess (i.e., one NEM per Cys residue). Under these conditions, the reaction slows as NEM is consumed and protein products are limited to those formed in the initial stages of the reaction. It is important to note that these conditions select for the most reactive Cys thiols and tend to exclude reactions at other thiols that might occur at higher reagent concentrations. Figure 1 shows HPLC elution profiles for the unreacted and NEM-modified protein after 0, 10, 30, and 60 min of reaction.

As the reaction progresses, the amount of eluted protein in the unmodified p7 peak diminishes and the amount of modified proteins eluting in peaks M-1, M-3, and M-6 increases. The amount of eluted protein in peak M-1 initially increases (from 0 to 30 min), then decreases as the reaction progresses (compare M-1 at 30 and 60 min). The amount of eluted protein in M-3 steadily increases over the course of the reaction, and the amount of protein eluting in peak M-6 is only detectable in the late stages of the reaction (Figure 1, 60 min). Protein eluting in peaks p7, M-1, M-3, and M-6 was collected and analyzed by MALDI-MS (Table 1). From these data we conclude that M-1 is modified by the addition of a single NEM moiety, M-3 is modified by the addition of three NEM moieties, and M-6 is modified by the addition of six NEM moieties. From the rate of appearance of M-1 (Figure 1, 10 min) and its apparent steady-state level at

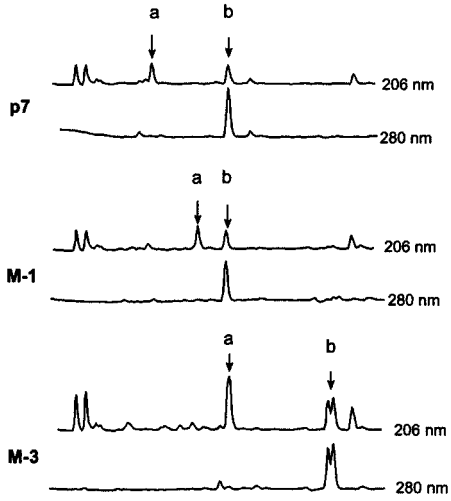

FIGURE 2: The samples eluting in p7, M-1, and M-3 (Figure 1) were lyophilized and then dissolved in 0.02 M tris-HCl, pH 7, and Arg-C endoproteinase (50:1 w/w NC protein−protease) was added and incubated for 1 h at 37 °C. Then chymotrypsin (TLCK treated) was added at the same ratio as Arg-C and then incubated for an additional 2 h at room temperature. After digestion all samples were analyzed by reversed-phase HPLC. Peaks were detected by UV absorption at 206 and 280 nm. Peak "a" on panel p7 was determined as KCGHQMKDCTER; on panel M-1 peak "a" consisted of one NEM-modified Cys residue, and it was KCGHQMKD(C-NEM)-TER; and on panel M-3 peak "a" was determined as K(C-NEM)-GHQMKD(C-NEM)TER. Peak "b" on panel p7 and M-1 was determined as GCW and on panel M-3 it was GC(-NEM)W.

intermediate reaction times (Figure 1, 30 min), we conclude that M-1 contains initial products of the reaction that are consumed in subsequent stages. Its ultimate decrease as the reactants are depleted (Figure 1, 60 min) suggests that it can undergo additional modifications to generate protein eluting in peak M-3. Taken together, these observations are consistent with a reaction pathway that proceeds from p7 to M-1, to M-3, and then to M-6.

Sequence Analysis of Initial Reaction Products. To determine whether M-1 and M-3 were discrete intermediates with uniform sites of modification and, if possible, to identify the sites of modification, each protein was digested with Arg C endoprotenase, followed by digestion with chymotrypsin to generate peptides suitable for peptide mapping and sequence analysis. The resulting peptide maps are shown in Figure 2 where relevant peptides are labeled "a" and "b". Peptide "b" contains the single Trp residue of p7 and was readily identified in each chromatogram by its distinguishing absorbance at 280 nm.

Comparing the peptide map for M-1 to p7 shows that peptide "a" in the p7 map moved to a new elution position in the M-1 map, while the elution positions of all other peptides remained unchanged. Peak "a" from M-1 was purified and analyzed by Edman degradation and was found to have the following sequence: KCGKEGHQMKD(C-NEM)TER corresponding to residues 38−52 of p7 (Figure 3). The measured mass of peptide "a" from M-1 was 1869.54 Da, consistent with the determined sequence and a single modified Cys residue (calculated 1873.78 Da). An Edman product of Cys-NEM was identified at step 12 of the sequence analysis, and no such product was seen in step 2. These data show that peak M-1 (Figure 1) contains a uniquely modified protein having a single Cys-NEM at position 49 in the sequence of NCp7. A comparison of the peptide map

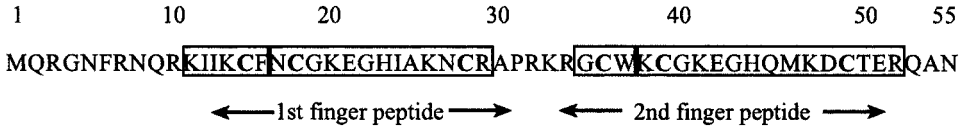

FIGURE 3: Amino acid sequence of HIV-1 MN NCp7. The resulting peptides corresponding to ZFs after protein was digested with ArgC and chymotripsin are shown in boxes.

of M-1 to the peptide map of M-3 shows that peptides "a" and "b" eluted later in the M-3 map while the elution of all other peptides remained unchanged. Peptides "a" and "b" from M-3 were analyzed by Edman degradation and mass spectrometry. Edman degradation of peptide "a" from M-3 showed Cys-NEM at steps 2 and 12 and the determined sequence was K(C-NEM)GKEGHQMKD(C-NEM)TER. Peptide "a" from M-3 had a mass of 2009.58 Da, consistent with the determined sequence and two modified Cys residues (calculated 1998.78 Da). Similar analysis showed that peptide "b" from M-3 had the sequence G(C-NEM)W corresponding to residues 48−50 in the sequence of NCp7. These data show that the protein purified in peak M-3 had modified Cys residues at positions 36, 39, and 49 of NCp7. The determined structures of M-1 and M-3 are consistent with a reaction pathway whereby modification initiates at residue Cys-49 in the second ZF of NCp7. The resulting modified p7 is stable enough to be detected as M-1; however, subsequent attack by NEM leads to a fully modified second ZF (M-3). Reactions on the Cys residues of the first ZF occur later in the reaction path.

Quenching of Trp-37 Fluorescence Does Not Coincide With Formation of Initial Products. NCp7 contains a single Trp residue (Trp-37) located in the first loop of the second ZF. The fluorescence response of Trp-37 in native NCp7 is quenched when the protein binds to nucleic acids or when zinc ions in the ZFs are removed (*16*). It was therefore of interest to determine the influence of the NEM modification reactions on the fluorescence of Tpr-37. Initial experiments conducted with protein and NEM concentrations identical to those used for the HPLC experiments described in Figure 1 showed no detectable change in Trp fluorescence over 1 h of observation. However, at higher ratios of NEM to protein (20:1 to 200:1), a decrease in Trp fluorescence over time was observed. Figure 4A shows the decrease in fluorescence for a reaction mixture containing a 200-fold excess of NEM over the NCp7. The overall magnitude of fluorescence quenching was comparable with that of a fully denatured apo-protein, and the time-dependent decrease in fluorescence followed pseudo-first-order kinetics (straight line on a semilog plot) with a $t_{0.5} = 9.7$ min. Figure 4B shows the results of analyzing the reaction mixture by HPLC at various times after the addition of NEM (0.5, 10, and 30 min). These results show that, under the conditions of a 200-fold excess of NEM, the $t_{0.5}$ for the disappearance of NCp7 and the appearance of M-3 is about 0.5 min (Figure 4B). However, there was little or no significant quenching of the Trp fluorescence associated with this phase of the reaction (Figure 4A). The $t_{0.5}$ for fluorescence quenching seems to correspond with the appearance of more fully modified products leading to M-6 (Figure 4B, 10 min). Thus we conclude that the quenching Trp-37 fluorescence does not accurately reflect the formation of the initial reaction products (M-1 and M-3) and that the reaction must progress beyond the M-3 intermediate before significant quenching occurs.

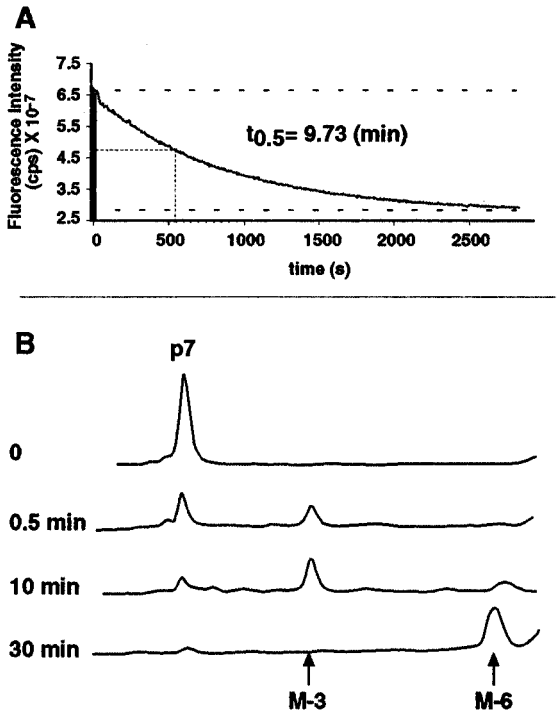

FIGURE 4: Fluorescence (A) and HPLC (B) analysis of NCp7 products of the reaction with NEM. (A) p7 coordinated with Zn^{2+} was dissolved in 20 mM Tris-HCl buffer (pH 7) at 2 μM concentration. NCp7 initial fluorescence was monitored for 2.5 min after which NEM at a protein−NEM ratio of 1:200 was added to the cuvette and its decrease in fluorescence was monitored over time. (B) p7 (2 μM) was mixed with a 200-fold excess of NEM in 20 mM Tris-HCl buffer (pH 7) and reacted for 0.5, 10, and 30 min. The products of the reaction were separated by reversed-phase HPLC.

To determine if the reaction pathway for p7 was dependent upon bound zinc, we added 1 mM EDTA to the protein before adding NEM and monitored the reaction as in Figure 1. After a reaction time of 3 min at 37 °C, modified protein peaks corresponding to M-1 and M-3 were absent and the protein was nearly quantitatively converted to M-6 (data not shown). These results show that the zinc-free protein is much more reactive than NCp7 and that the unique reaction path for NCp7 is dependent upon coordinated zinc and the native structure of NCp7.

Reaction of First and Second ZF Peptides With NEM. In the context of the whole protein, the second ZF is much more reactive than the first ZF. It was, therefore, of interest to determine if the apparent preferential susceptibility of the second ZF is an intrinsic attribute of the peptide metal ion complex. It was previously shown that 18 residue peptides corresponding to the first and second ZFs of HIV-1 NCp7 bind zinc and have the same side chain orientations as those in the whole protein (*17*). Synthetic peptides corresponding to residues 13−30 and 34−51 in the whole protein (Figure 3) were complexed with zinc as previously described (*17*) to generate the first and second ZF peptides. An equal molar

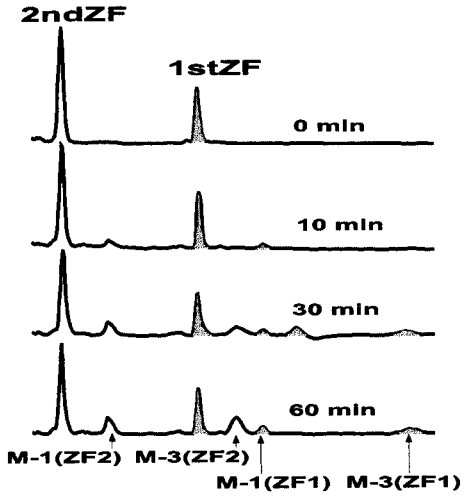

FIGURE 5: An equal molar mixture of zinc-complexed synthetic peptides (7.5 μM) corresponding to residues 13−30 and 34−51 in the whole protein was reacted with NEM (1:1 ratio of Cys residue to NEM), and the reaction products were separated by HPLC. Peaks were detected by UV absorption at 206 and 280 nm and analyzed by sequencing. Shaded peaks correspond to the unreacted first finger and its reaction products. All peaks were analyzed by MALDI-TOF MS to determine molecular masses of intermediate and fully modified peptides.

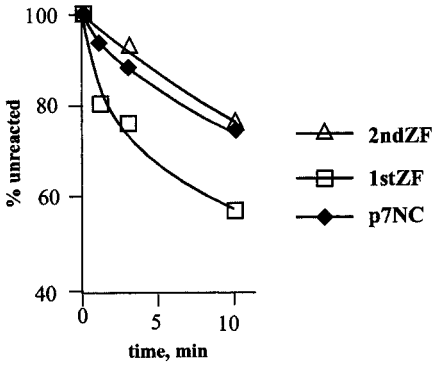

FIGURE 6: Summary of the results of the chromatography profiles (Figures 1 and 5) analyzed using peak areas to determine the amounts of NCp7 and ZF peptides before and after reaction.

mixture of these peptides was treated with NEM (1:1 ratio of Cys residues to NEM), and the reaction was analyzed by HPLC as before. Modified forms of the second ZF peptide were distinguishable on the basis of tryptophan absorption at 280 nm. The resulting chromatograms are shown in Figure 5 and were analyzed to determine the area of each peak as before.

The determined areas were plotted to obtain the apparent reaction rates for the individual peptides and then compared to the whole protein (Figure 6). These data show that, as independent peptides, both the first and second ZFs are reactive with NEM and that the first ZF peptide is more reactive than the second ZF peptide. The data suggest that the preferential reactivity of the second ZF of NCp7 is due to an apparent sequestering of the first ZF in the context of the whole protein. The initial reaction product for the second ZF peptide (M-1, in Figure 5) was analyzed by mass spectrometry and Edman degradation as before to determine if it was uniquely modified at Cys-49, as was seen for the whole protein. These data show that the modified peptide contained one Cys-NEM residue (2252.64 Da; calculated, 2193.4 Da) but the modification was distributed over all three

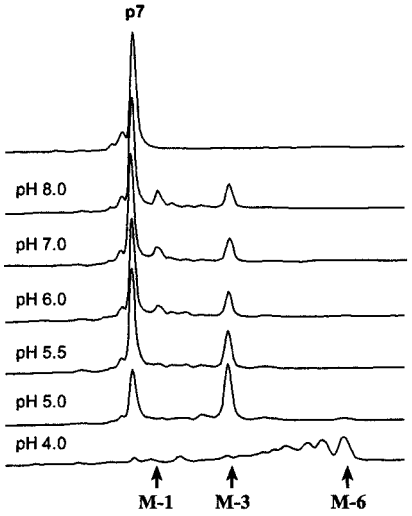

FIGURE 7: pH dependency of NEM modification of HIV-1 NCp7. The protein, 2 μg per analysis, at 7.5 μM was incubated with a 6-fold excess of NEM in 0.02 M $NaCH_3COOH$, pH 8.0, 7.0, 6.0, 5.5, 5.0, and 4.0 at 37 °C. After the reaction for 30 min, samples were separated by reversed-phase HPLC.

Cys residues. We conclude that selective modification of Cys-49 in NCp7 is due to the sequestering of other Cys residues by the native conformation of the protein.

pH Dependency of NEM Modification of HIV-1 NCp7. To investigate the pH dependency for the reaction, we treated NCp7 with NEM at various pHs (8, 7, 6, 5.5, 5, and 4) and analyzed it by using HPLC as before (Figure 7). These results showed that the reaction rate and HPLC profile remained constant between pH 6 and 8. The reaction rate was slightly faster at pH 5.5: the amount of M-1 was decreased and the amount of M-3 was increased. This trend was even more pronounced at pH 5.0. At pH 4, the reaction rate was much faster (note the absence of p7, M-1, and M-3) and most of the proteins eluted as though they were products produced by a reaction with the apo-protein. We take these results to suggest that the NCp7 protein is stable between pH 8 and ~5.5 and that between pH 5 and 4 the protein is denatured and the bound zinc ions are released. The apparent reduction in the amounts of M-1 detected after 10 min reaction time at pHs 6.0, 5.5, and 5.0 (Figure 7) would support the suggestion that the M-1 intermediate is more sensitive to acid than either NCp7 or M-3.

NEM Modification of HIV-1 NCp7 Bound to Short Oligonucleotides. The NEM modification reaction was extended to probe the architecture of NCp7 bound, to short oligodeoxynucleotides. Several short single-stranded DNA nucleotides, $d(T)_8$, $d(G)_8$, $d(GT)_4$, $d(GT)_2$, and $d(IT)_4$ were selected for this study on the basis of their known high-affinity interactions with p7NC (*16*). These octameric nucleotides form 1:1 complexes with NCp7, and at low salt concentrations the binding is essentially quantitative. The experiments here were designed to achieve stoichiometric binding of NCp7 by adding the nucleotides in a 2-fold molar excess over the protein in low-ionic-strength buffer. These conditions ensure that any observed reactions with NEM can be attributed to modifications of the protein in the bound complex and not to free protein. As before, reactions were initiated with the addition of NEM and stopped by dilution and separation by HPLC. Figure 8 shows the HPLC elution

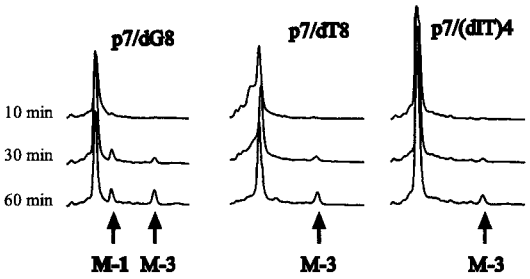

FIGURE 8: Zn^{2+}−NCp7 (7.5 μM) was incubated with single-stranded DNA nucleotides, $d(G)_8$, $d(T)_8$, and $d(IT)_4$, at a protein−oligonucleotide ratio of 1:2 at 37 °C in 0.02 M Tris-HCl, pH 7.0, for 5 min, followed by NEM addition. The reaction products were separated after 10, 30, and 60 min by reversed-phase HPLC. Peaks were detected at 206, 260, and 280 nm.

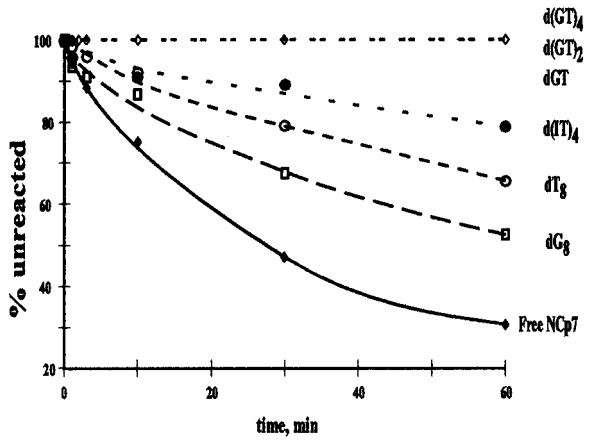

FIGURE 9: Reaction rates of free and oligonucleotide-bound NCp7 with NEM, and a summary of the results of the chromatography profiles (Figures 1 and 8) analyzed using peaks areas and to determine the amounts of NCp7 before and after reaction.

profiles obtained after reacting NCp7−oligonucleotide complexes with NEM for varying times. Reaction rates were estimated by the decrease in unmodified protein with time and are plotted in Figure 9.

In all cases, the NCp7−oligonucleotide complexes were less reactive with NEM than the free protein. However, the rates of reactions for the NCp7−oligonucleotide complexes were dependent upon the bound nucleotide sequence. When bound to dG_8, NCp7 was more reactive than when bound to dT_8 or $d(IT)_4$, and the NCp7 bound to $d(GT)_4$ did not react under these conditions. These data show that molecular features governing the reactivity of the protein with NEM are not identical in the free and bound state and vary depending upon the base sequence of the interacting oligonucleotide. This conclusion is reinforced by the observations that the distribution of initial reaction products (M-1 and M-3) is also dependent upon the nature of the bound oligonucleotide. All three of the reactive oligonucleotide−protein complexes react to form the M-3 modified protein. However, the transient intermediate M-1 is only apparent in the analysis of the reaction products of NCp7−dG_8 (Figure 6) and is either absent or barely discernible in the analysis of reaction products for the $d(IT)_4$ and dT_8 complexes.

We previously showed that $d(TG)_2$ could form a stable complex with NCp7; however in this case one protein bound two short oligonucleotides (i.e., NCp7−2[$d(TG)_2$]) (*16*). It was of interest to determine whether the complex with $d(TG)_2$ would exhibit the same resistance to NEM as the

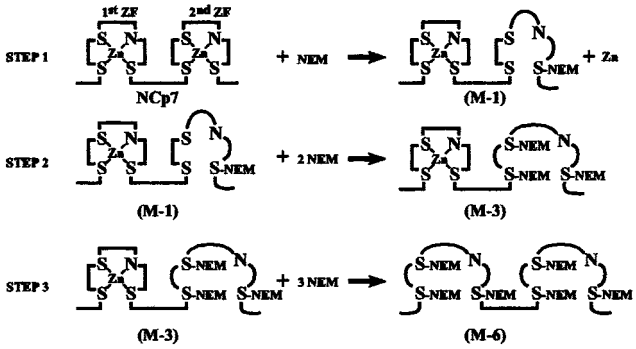

FIGURE 10: Schematic representation of the NCp7 reaction pathway with NEM.

complex with $d(TG)_4$. NCp7 was mixed with a 4-fold excess of $d(TG)_2$ in a low-salt buffer, treated with NEM, and analyzed by HPLC as before. These experiments showed that, over the course of 1 h reaction time, the NCp7−2$d(TG)_2$ complex was inert to NEM modification (data not shown). Similar experiments using d(TG) showed that even the dinucleotide could protect NCp7 from the reaction with NEM, although higher ratios of nucleotide to protein were required.

DISCUSSION

NEM Initially Reacts with Cys-49 in the Second ZF of HIV-1 NC Protein. HIV-1 NCp7 was treated with NEM at a ratio of one NEM per cysteine thiol (6 NEM/NC protein), and the reaction was followed by analyzing the amounts of unreacted protein and reaction products at increasing reaction times. The unreacted and modified NC proteins were separated by HPLC, and the initial reaction product (M-1) was isolated as a single peak of purified protein well-separated from the unreacted protein. Analysis of the initial reaction product by MS and sequence analysis showed that it contained one NEM-Cys residue located at position 49 in the sequence of NCp7. Time studies showed that the initial product (M-1) accumulated in the reaction mixture as a transient intermediate in the reaction pathway and subsequently all three Cys residues in the second ZF were modified (M-3). The reaction then proceeded to modification of Cys residues in the first ZF to generate the final product (M-6). The determined reaction path is depicted diagrammatically in Figure 10. Our data is completely compatible with previous observations showing that the second ZF of NCp7 is more reactive with disulfide reagents than the first ZF (*18, 19*) and extends to identification of the initial site of attack.

The Reaction Pathway Initiating at Cys-49 is Determined by NC Conformation. To determine if the sequential order of reaction was dependent upon the zinc-bound state of the protein, we treated the zinc-free protein with NEM and analyzed it as described above. These results showed that the apo-protein reacted much faster with NEM and did not follow a discrete reaction pathway. To determine if the reaction path was strictly governed by zinc coordination of the independent ZFs, we complexed synthetic peptides (18 residues each) containing the metal-binding sequences of the first and second ZFs with zinc and reacted then with NEM. Analysis of the reaction products showed that, under identical conditions, the isolated first ZF peptide reacted faster than the second ZF peptide and the native NCp7 protein. Also,

the sequence analysis showed that the reactions of ZF peptides with NEM did not follow a discrete reaction path. Thus, the reaction path for the NCp7 protein cannot be explained by the reaction rates and pathways exhibited by the isolated finger peptides. The data strongly suggest that zinc coordination imparts a three-dimensional structure to the NC protein which exposes Cys-49 to the initial reaction with NEM while sheltering all other Cys thioles.

Implication for NCp7 Conformational Flexibility in Solution. The three-dimensional conformation of the ZFs of HIV-1 has been determined by NMR (*20*). However, the N-terminal peptide region, the linker region, and the C-terminal region do not yield sufficient NMR data to determine conformation. NMR analysis has suggested an interaction between the first and second ZFs (*21*−*23*); however, the observed NMR signals for this interaction are relatively weak, so the significance of the observation is controversial. With the present knowledge, one could speculate that many conformational states are possible, depending upon the flexibility of the protein in the N-terminal, linker, and C-terminal peptide segments. The results of NEM modification suggest that, in the NC protein, the ZFs do not react as though they are free in solution. One possible explanation suggested by the NMR results of Morellet et al. (*22*) is that conformational flexing in the linker peptide may allow conformers with close approach of the two ZFs. Our own computational modeling of the protein suggests that conformers ranging from the fully extended linker region to more collapsed structures that bring the two ZFs into close proximity are possible without introducing strain. These conformers are compatible with the available NMR data for the solution structure of the protein (*21*−*23*) and are achieved through allowable energy-minimized orientations of the peptide segment linking the two ZF domains. In more fully extended conformers Cys thiols in the first and second ZFs can be exposed to the solvent depending upon local side chain orientations. In the intermediate and collapsed conformers, residues in the linker peptide and in the ZFs can restrict the solvent exposure of Cys-18, Cys-28, and Cys-36. In all conformers, solvent access to Cys-15 in the first ZF is restricted by the N-terminal peptide and solvent access to Cys-39 and Cys-49 in the second ZF is controlled by the orientation of the C-terminal peptide. While the data here are not offered as proof, we suggest that the NC protein in solution may occupy a family of closely related conformers that favor a partly collapsed orientation (induced by flexible bending in the linker region) which restricts access to Cys-28 and Cys-36 and with the C-terminal peptide restricting access to Cys-39 but allowing access to Cys-49.

M-1 May Retain Significant Conformation. The finding that the first and second ZF peptides are reactive with NEM and that the second ZF can be attacked at all three Cys residues has implications for the nature of the transient intermediate M-1 in the reaction path for native NCp7. The observation suggests that M-1 still retains enough native conformation to retard reactions at Cys-36 and Cys-39. If M-1 immediately lost all nativelike conformation in the second ZF region, we might reasonably expect Cys residues 36 and 39 to react rapidly with NEM (as for apo-protein) and the steady-state level of M-1 would be very small. Thus we conclude that M-1 retains a conformation which retards the reactivity of Cys residues 36 and 39 with NEM.

Fluorescence of Trp-37 and NEM-Modified Forms of NCp7. Our observation that the fluorescence of Trp-37 (located in the second ZF) is not significantly quenched until the reaction has progressed beyond the M-3 intermediate (modification of all three Cys residues in the second ZF) suggests that intermediates M-1 and M-3 retain some nativelike conformation. We assume that zinc remains bound to the first ZF of M-1 and M-3 and that the NEM-modified Cys thiols of M-3 do not bind zinc ions. However, the zinc binding properties of M-1 are unclear. It is interesting to note that reaction mixtures that do not progress beyond the formation of M-3 (ratios of 6:1, NEM−NCp7) show no appreciable decrease in fluorescence over 1 h at pH 7, suggesting stable conformational states. It cannot be ruled out that these conformers represent kinetically trapped intermediates, but their stability suggests thermodynamically stable states. The stability of M-1 is markedly reduced at pH 6.0 and below (Figure 7), suggesting that ionization of a His residue might be a destabilizing factor. This observation is compatible with the suggestion that M-1 may still retain bound zinc in the modified second ZF.

Influence of Bound Nucleotide on the Reaction with NEM. We previously reported that NCp7 forms 1:1 complexes with short single-stranded nucleotides (*16*). To explore the effect of nucleotide binding, we complexed NC protein with short oligonucleotides (eight bases) under conditions of low ionic strength and analyzed the reaction with NEM. The results show that the bound nucleotide sequence influences the reactivity of the protein. The observed order of reactivity with NEM is NCp7 > NCp7−$d(G)_8$ > NCp7−$d(T)_8$ > NCp7−$d(TI)_4$ ≫ NCp7−$d(GT)_4$, which does not correlate with the observed order of binding strengths ($d(GT)_4$ > $d(G)_8$ > $d(T)_8$) (*16*). The oligonucleotide with the highest binding strength ($d(GT)_4$) protects NC from reaction with NEM, whereas the oligonucleotide with the next highest binding strength ($d(G)_8$) affords little or no protection. These observations suggest that conformational differences induced by the bound nucleotide sequence may influence the accessibility of Cys-49 to reaction with NEM. Induced conformational changes that restrict access to Cys-49 or alter the reaction path should also result in slower rates of formation and lower steady-state levels for M-1. This hypothesis may explain apparent differences in the reaction paths induced by nucleotide binding. NCp7−$d(G)_8$ reacts with NEM at a rate that is only slightly slower than the rate for free NCp7 (Figure 9), and the steady-state levels of M-1 are comparable for the two reactions. However, NCp7 bound to $d(T)_8$ or $d(TI)_4$ reacts significantly more slowly than the free protein (Figure 9) and also shows much lower steady-state levels of M-1 (Figure 8). It is possible that the reactions of NEM with NCp7 bound to $d(T)_8$ or $d(TI)_4$ initiate at Cys-49, but the rates of formation of M-1 are slower than for the free protein. However, it is also possible that the conformations of the nucleotide-bound protein promote an initial reaction at Cys residues other than Cys-49. In the case of NCp7 bound to $d(TG)_4$, the induced protein conformation is such that NEM access to Cys residues is highly restricted.

We previously reported that $d(TG)_4$ binds to NCp7 and forms a more stable complex than NCp7 bound $d(IT)_4$ (*16*). In both cases the protein interacted with five nucleotide bases

and the two complexes differed only in the purine base. These data suggest that, in the bound state, the purine base G might participate in additional stabilizing interactions that could include H-bonding. In this regard it is interesting to note that the recent three-dimensional structure of NCp7 bound to a stem-loop RNA template shows intimate interactions between purine bases (Gs) and the ZF domains of the bound protein (*24*). The data presented here show that, in addition to being more stable, the complex with $d(TG)_4$ is not reactive with NEM under conditions where the complex with $d(IT)_4$ is reactive (Figure 9). This strongly suggests that the protein conformations are not identical in the two complexes and that the difference is due to interactions with the purine base.

These types of studies should provide important insights regarding NC interactions with nucleic acid within the virus. It is currently proposed that all 2000−2500 molecules of NC protein are bound to the viral RNA. This suggests that the protein interacts with a number of different RNA sequences and that these interactions are likely to involve various binding strengths. Knowledge of the protein topography assists in predicting the NC involvement in biological processes such as genomic RNA packaging and early infectious events. The data reported in this paper suggest that assays designed to detect anti-NC compounds should include NC−oligonucleotide complexes.

The NEM reaction pathway with the zinc-coordinated NC protein shows that NCp7 exists in a conformation that cannot be duplicated by free ZF peptides. Oligonucleotide binding to NC protein decreases or prevents reactivity with NEM, suggesting that the bound oligonucleotide shields the reactive cysteine residues and/or indicates conformational changes in the protein that make the cysteine residues unavailable for reaction with NEM.

ACKNOWLEDGMENT

We are grateful to Mrs. Laura K. Busch for contribution in the expression of NCp7, which was a part of her master thesis, and to Dr. Jose Casas-Finet for the fruitful discussion and suggestions.

REFERENCES

1. Henderson, L. E., Sowder, R., Copeland, T. D., Smythers, G., and Oroszlan, S. (1984) *J. Virol. 52*, 492−500.
2. Copeland, T. D., Morgan, M. A., and Oroszlan, S. (1984) *Virology 133*, 137−145.
3. Aldovini, A., and Young, R. A. (1990) *J. Virol. 64*, 1920−1926.
4. Dupraz, P., Oertle, S., Meric, C., Damay, P., and Spahr, P. F. (1990) *J. Virol. 64*, 4978−4987.
5. Gorelick, R. J., Henderson, L. E., Hanser, J. P., and Rein, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8420−8424.
6. Gorelick, R. J., Nigida, S. M., Jr., Bess, J. W., Jr., Arthur, L. O., Henderson, L. E., and Rein, A. (1990) *J. Virol. 64*, 3207−3211.
7. Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1993) *J. Virol. 67*, 4027−4036.
8. Schwartz, M. D., Fiore, D., and Panganiban, A. T. (1997) *J. Virol. 71*, 9295−9305.
9. Zhang, Y., and Barklis, E. (1995) *J. Virol. 69*, 5716−5722.
10. Gorelick, R. J., Chabot, D. J., Ott, D. E., Gagliardi, T. D., Rein, A., Henderson, L. E., and Arthur, L. O. (1996) *J. Virol. 70*, 2593−2597.
11. Gallay, P., Swingler, S., Song, J., Bushman, F., and Trono, D. (1995) *Cell 83*, 569−576.
12. Henderson, L. E., Rice, W. G., and Arthur, L. O. (1995) in U.S. Patent Application USSN 08/312, pp 331, U.S.A.
13. 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., and Kun, E. (1993) *Nature 361*, 473−475.
14. Rice, W. G., Supko, J. G., Malspeis, L., Buckheit, R. W., Jr., Clanton, D., Bu, M., Graham, L., Schaeffer, C. A., Turpin, J. A., Domagala, J., Rocco, G., Bader, J. P., Halliday, S. M., Coren, L., Sowder, R. C., II, Arthur, L. O., and Henderson, L. E. (1995) *Science 270*, 1194−1197.
15. Wu, W., Henderson, L. E., Copeland, T. D., Gorelick, R. J., Bosche, W. J., Rein, A., and Levin, J. G. (1996) *J. Virol. 70*, 7132−7142.
16. Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E. (1998) *J. Virol. 72*, 1902−1909.
17. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry 29*, 329−340.
18. Hathout, Y., Fabris, D., Han, M. S., Sowder, R. C., 2nd, Henderson, L. E., and Fenselau, C. (1996) *Drug. Metab. Dispos. 24*, 1395−1400.
19. Tummino, P. J., Scholten, J. D., Harvey, P. J., Holler, T. P., Maloney, L., Gogliotti, R., Domagala, J., and Hupe, D. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 969−973.
20. Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., Jr., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C. I., Hare, D. R., and Arthur, L. O. (1992) *Protein Sci. 1*, 563−574.
21. Mely, Y., Jullian, N., Morellet, N., De Rocquigny, H., Dong, C. Z., Piemont, E., Roques, B. P., and Gerard, D. (1994) *Biochemistry 33*, 12085−12091.
22. Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J. L., Fournie-Zaluski, M. C., and Roques, B. P. (1994) *J. Mol. Biol. 235*, 287−301.
23. Lee, B. L., De Guzman, R. N., Turner, B., Tjandra, N., Bax, A., and Summers, M. F. (1998) *J. Mol. Biol. 279*, 633−649.
24. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−388.

BI980907Y