



Insights into the Regioselectivity and RNA-binding Affinity of HIV-1 Nucleocapsid Protein from Linear-scaling Quantum Methods

Jana Khandogin, Karin Musier-Forsyth and Darrin M. York*

*Department of Chemistry
University of Minnesota
Minneapolis, MN 55455, USA*

Human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein (NC) plays several important roles in the viral life-cycle and presents an attractive target for rational drug design. Here, the macromolecular reactivity of NC and its binding to RNA is characterized through determination of electrostatic and chemical descriptors derived from linear-scaling quantum calculations in solution. The computational results offer a rationale for the experimentally observed susceptibility of the Cys49 thiolate toward small-molecule electrophilic agents, and support the recently proposed stepwise protonation mechanism of the C-terminal Zn-coordination complex. The distinctive binding mode of NC to SL2 and SL3 stem-loops of the HIV-1 genomic RNA packaging signal is studied on the basis of protein side-chain contributions to the electrostatic binding energies. These results indicate the importance of several basic residues in the 3₁₀ helical region and the N-terminal zinc finger, and rationalize the presence of several evolutionarily conserved residues in NC. The combined reactivity and RNA-binding study provides new insights that may contribute toward the structure-based design of anti-HIV therapies.

© 2003 Elsevier Ltd. All rights reserved

Keywords: nucleocapsid protein; chemical reactivity; RNA binding; pK_a; linear-scaling quantum

*Corresponding author

Introduction

The human immunodeficiency virus type 1 (HIV-1) RNA genome encodes the Gag polyprotein, which is processed into several mature proteins, including the matrix protein, the capsid protein and the nucleocapsid protein (NC), also known as NCp7. The mature NC protein acts as a nucleic acid “chaperone” during reverse transcription^{1–5} and facilitates various nucleic acid rearrangements, including annealing of the tRNA primer to the viral RNA template,^{3,6–8} and minus and plus-strand transfer.^{9–14} NC, a highly basic protein, contains two copies of the CCHC-type Zn²⁺-coordinating peptide sequence, Cys-X₂-Cys-X₄-His-X₄-Cys (X = variable amino acid residue), that are separated by a short flexible linker sequence (see Figure 3). All known ortho-retroviruses contain one or two copies of the

CCHC-type Zn²⁺ finger domains,^{3,15} also referred to as zinc “knuckle” motifs. Recently, it has been shown that NC’s zinc finger motifs are required for optimal nucleic acid chaperone activity.^{16–19} The high-resolution NMR structures of HIV-1 NC bound to SL2 and SL3 stem-loops of viral RNA reveal specific interactions between both zinc fingers and the RNA bases.^{20,21} The distinctive three-dimensional structure of NC’s zinc fingers, their critical role in the viral life-cycle,^{22–27} and their mutationally non-permissive nature have made them attractive targets for development of antiviral drug therapies.^{28–33}

A series of organic compounds have been identified that can chemically modify the Zn²⁺-binding cysteine residues resulting in ejection of zinc ions, disruption of the protein structure and ultimately loss of viral infectivity.^{31,34–36} Several experiments employing fluorescence spectroscopy and electrospray ionization mass spectrometry revealed that the zinc ion ejection occurs at the carboxyl terminal zinc finger first, and that Cys49 is the target of the initial electrophilic attack.^{37–39} Furthermore, the reaction rate of NC toward electrophilic attack

Abbreviations used: HIV-1, human immunodeficiency virus type 1; H(L)OMO, highest(lowest)-lying occupied molecular orbitals; NC, nucleocapsid protein.

E-mail address of the corresponding author:
york@chem.umn.edu

was observed to be reduced upon its binding to oligonucleotides.³⁸

Computational studies performed to date have been directed toward understanding the reaction mechanism for electrophilic attack on NC's zinc fingers and their differential reactivity. Using small model structures for the isolated zinc fingers, high-level quantum chemical calculations (Hartree-Fock and density-functional theory) have been carried out to study the reaction path,^{40,41} and the relative chemical reactivity of the zinc finger thiolate groups.⁴² A protein docking study has also been performed with a variety of small electrophiles to evaluate the steric accessibility of both zinc fingers.³⁰ Recently, the Poisson-Boltzmann method with fixed force field charges was used to investigate the protein packing density and classical electrostatic screening of the Zn²⁺-binding atoms in different classes of zinc fingers.⁴³ In this study, the relative chemical reactivity of the zinc fingers, as well as the individual thiolate groups was attributed solely to the electrostatic screening due to the protein environment. While this approach is fast, it neglects the explicit electronic and solvent-induced polarization, as well as the contribution due to atomic multipoles. Although the previous quantum studies shed light on the chemical reactivity of NC's zinc fingers, it would be desirable to perform a quantum chemical study of this system that simultaneously takes into account the effect of the full protein environment, solvation, and nucleic acid binding. It is an interesting question as to what degree a "supermolecular" approach is necessary to quantify chemical reactivity indices associated with the NC zinc knuckles. Since the C and N-terminal zinc coordination domains are in moderately close proximity to one another, it is likely that they have a mutual polarization effect. As indicated by density-functional calculations⁴² in the absence of solvation, the relative reactivity of the zinc-coordinating thiolate groups is sensitive to the size of the zinc domain model. In addition to this size effect, spatial differences in the solvated macromolecular electrostatic potential of both the unbound and RNA-bound forms (Figures 1 and 2), are expected to influence the relative values of the reactivity indices. These observations are suggestive that the global protein and protein/RNA environment may play an important role in the relative reactivity of the NC zinc knuckles; however, a systematic study was not made here.

The present work focuses on the quantum characterization of the NC-binding domain in the fully solvated protein environment, in both unbound and RNA-bound forms, using linear-scaling electronic structure methods.^{44,45} Linear-scaling electronic structure methods allow the determination of a host of chemical properties that greatly extend the scope of information obtainable from purely classical models.⁴⁶ These include response properties and chemical reactivity indices that take into account quantum many-body effects and provide an accurate description of the electronic structure that governs biomolecular interactions. In this work, the quantum

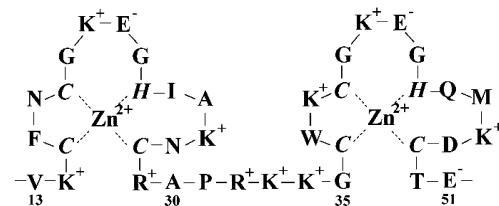


Figure 1. Amino acid sequence of zinc knuckles in isolated HIV-1 NC. The N-terminal (ZF1) and C-terminal (ZF2) zinc-binding sequences span residues 13–30 and 35–51, respectively. The Zn²⁺-coordinating residues are in italics. Zinc and charged residues are indicated with superscript + (positive) or - (negative).

chemical descriptors applied to NC protein can be divided into two classes: (1) descriptors used to characterize electrostatic properties such as solvent-polarized atomic charges, molecular electrostatic potential maps, and proton potentials; and (2) descriptors used to characterize chemical reactivity such as atomic softness indices and local hardness maps. The latter are related to the chemical softness and hardness of inorganic chemistry and have relation to the frontier orbital theory pioneered by Fukui (see Methods for additional details).

The objective of the present study is to evaluate the regioselective reactivity of HIV-1 NC and to explore the electrostatic binding mode of NC to SL2 and SL3 stem-loops, which are part of the viral RNA genome packaging signal. Toward this end, the electrostatic features of NC's zinc finger regions are evaluated by comparing the solvent-polarized molecular electrostatic surface potential at both finger regions and the solvent-polarized atomic charges of Zn²⁺-binding atoms. The relative chemical reactivity of the zinc fingers and the Zn²⁺-coordinating thiolate groups are assessed by the local hardness map and atomic softness indices. A qualitative prediction of relative acidity of the zinc finger thiols in the NC apoprotein is given by comparing their relative proton potentials, which have been shown to correlate with experimental pK_a values (J.K. & D.M.Y., unpublished results).⁴⁷

These results offer a theoretical basis for the regioselectivity of the NC zinc finger domains toward electrophilic attack. Finally, individual residue contributions to the electrostatic binding of NC to the SL2 and SL3 stem-loops are analyzed and compared to mutational data to better understand the origin of the NC–RNA interaction. Taken together, these results offer new insights that may contribute toward the rational design of anti-NC HIV therapies.

Results and Discussion

Molecular electrostatic properties

The electrostatic potential, $\phi(\vec{r})$, is a measure of the degree of stabilization of a unit test charge at the position \vec{r} . A particularly useful descriptor for

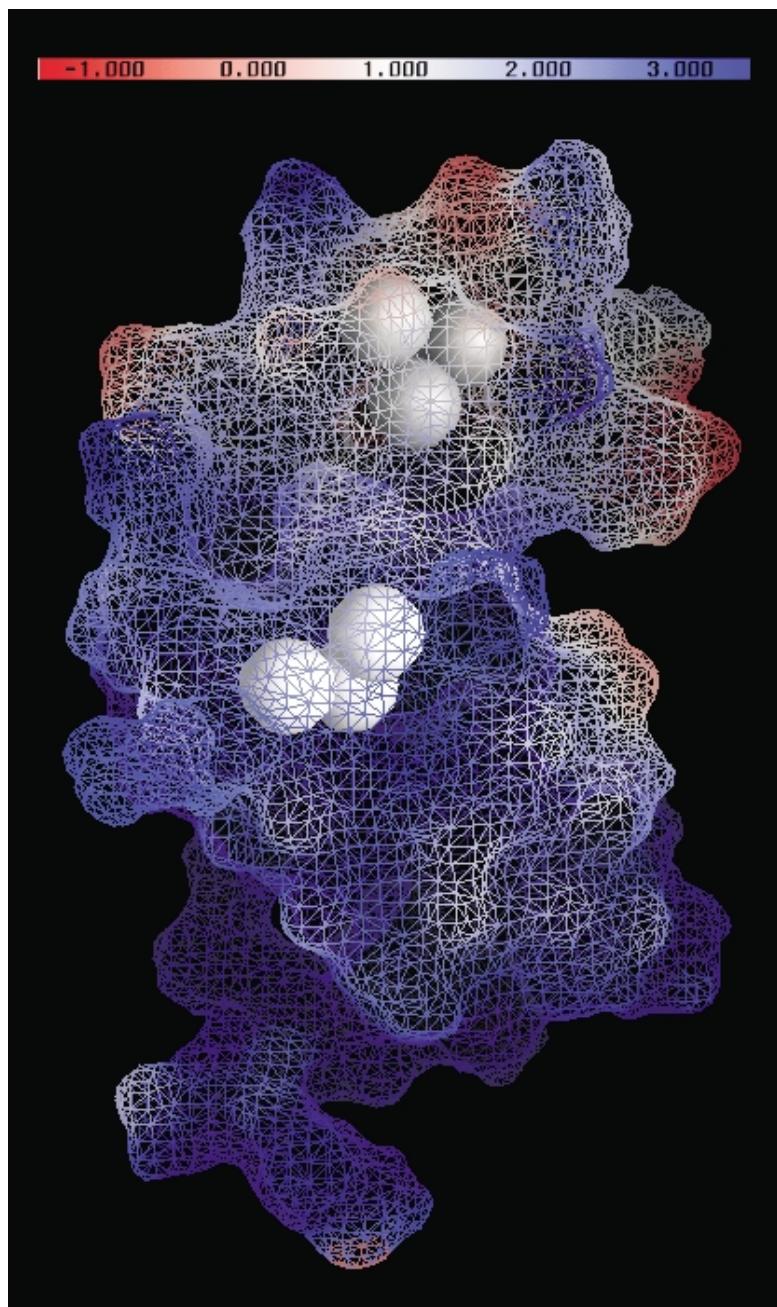


Figure 2. The electrostatic potential map of the NC-SL3 RNA complex. The surface is rotated to give the best view of the solvent-exposed Zn^{2+} -coordinating sulfur atoms (shown as white spheres). The N-terminal zinc finger (ZF1) is at the bottom of the Figure, while the C-terminal finger (ZF2) is at the top. The electrostatic potential is given in units of kT/e ($T = 298$ K), and the color scale is adjusted to allow the best view of the difference electrostatic potential on zinc knuckle regions.

biological macromolecules is a map of $\phi(\vec{r})$ on the solvent-accessible (or solvent-excluded) molecular surface, where interactions with molecules from the environment are likely to take place.^{47–49} Such an “electrostatic potential map” can be readily visualized and used to probe the electrostatic features over the binding interface. Figures 1 and 2 show the electrostatic potential maps of unbound NC and the NC-SL3 RNA complex, respectively. While in the unbound protein, the surface around the N-terminal zinc finger (ZF1) clearly displays

more positive electrostatic potential (indicated in blue) than that around the C-terminal zinc finger (ZF2), the difference is much smaller in the protein–RNA complex. Following the designations used by De Guzman *et al.*,²⁰ ZF1 (Val13-Ala30) contains four basic and one acidic residues while ZF2 (Gly35-Glu51) has three basic and three acidic residues (Figure 3). Taking into account the charges on cysteine and the Zn^{2+} , the overall formal net charge of ZF1 is +2 and that of ZF2 is −1 (Figure 3). The more positive electrostatic potential around the

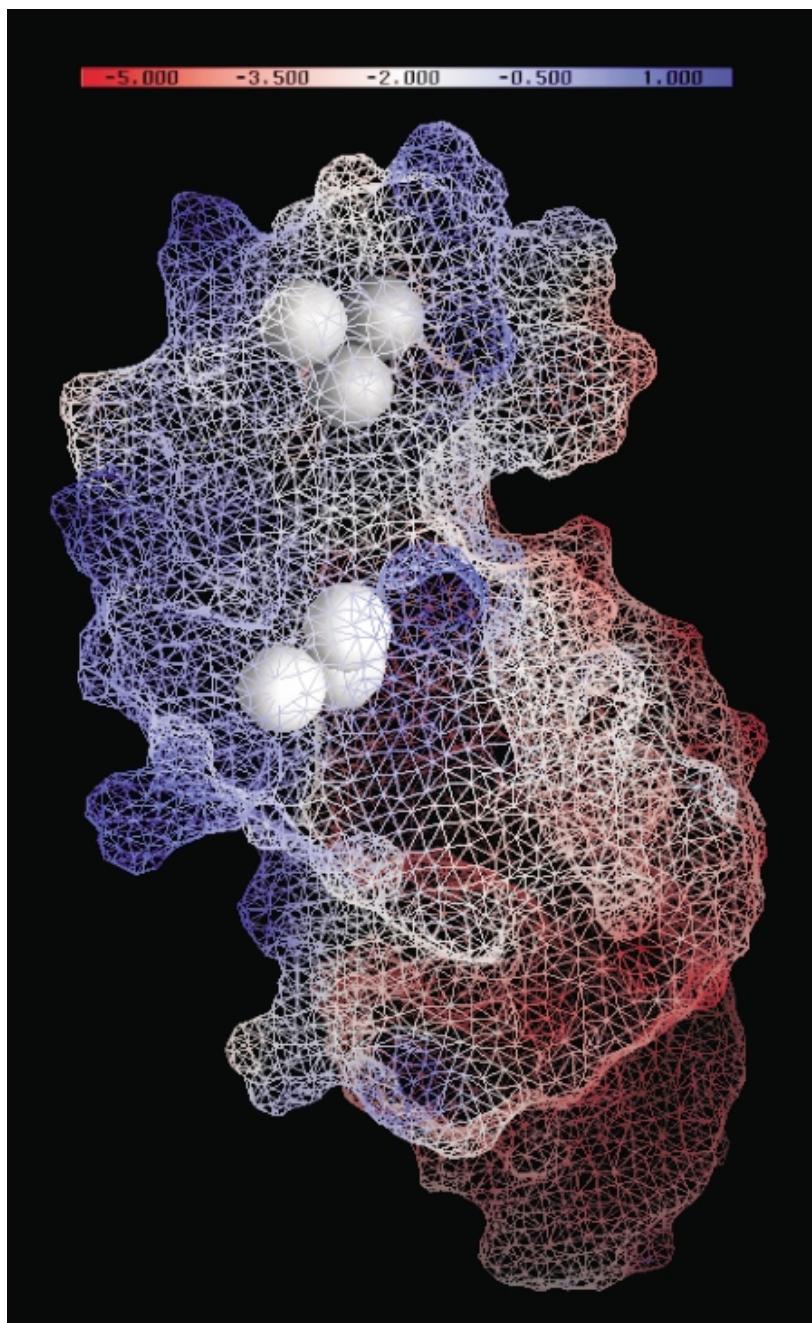


Figure 3. The electrostatic potential map of NC. The surface is rotated to give the best view of the solvent-exposed Zn^{2+} -coordinating sulfur atoms (shown as white spheres). The N-terminal zinc finger (ZF1) is at the bottom of the figure, while the C-terminal finger (ZF2) is at the top. The electrostatic potential is given in units of kT/e (k = Boltzmann constant, e = atomic unit of charge, and $T = 298 \text{ K}$), and the color scale is adjusted to allow the best view of the difference electrostatic potential on zinc knuckle regions.

ZF1 region reflects this difference in the total net charge. Upon RNA binding, the electrostatic interaction of the basic residues of NC with the RNA backbone and bases serves to decrease the positive surface potential on NC. Since the N terminus has more electrostatic contacts with SL3 RNA (see later discussions on binding energies), the electrostatic potential in the ZF1 region is lowered more upon RNA binding, resulting in similar electrostatic potential for the two zinc-binding regions in

the complex. This effect is also seen for NC binding to SL2 RNA (data not shown).

Quantum mechanically derived solvent-polarized atomic charges provide useful information about the distribution of electrons in a solvated biological macromolecule. Table 1 gives the average atomic charges and RMS deviations due to conformational variations on the Zn^{2+} -coordinating atoms in the NC and NC-SL3 RNA complex. The atomic charge differences between

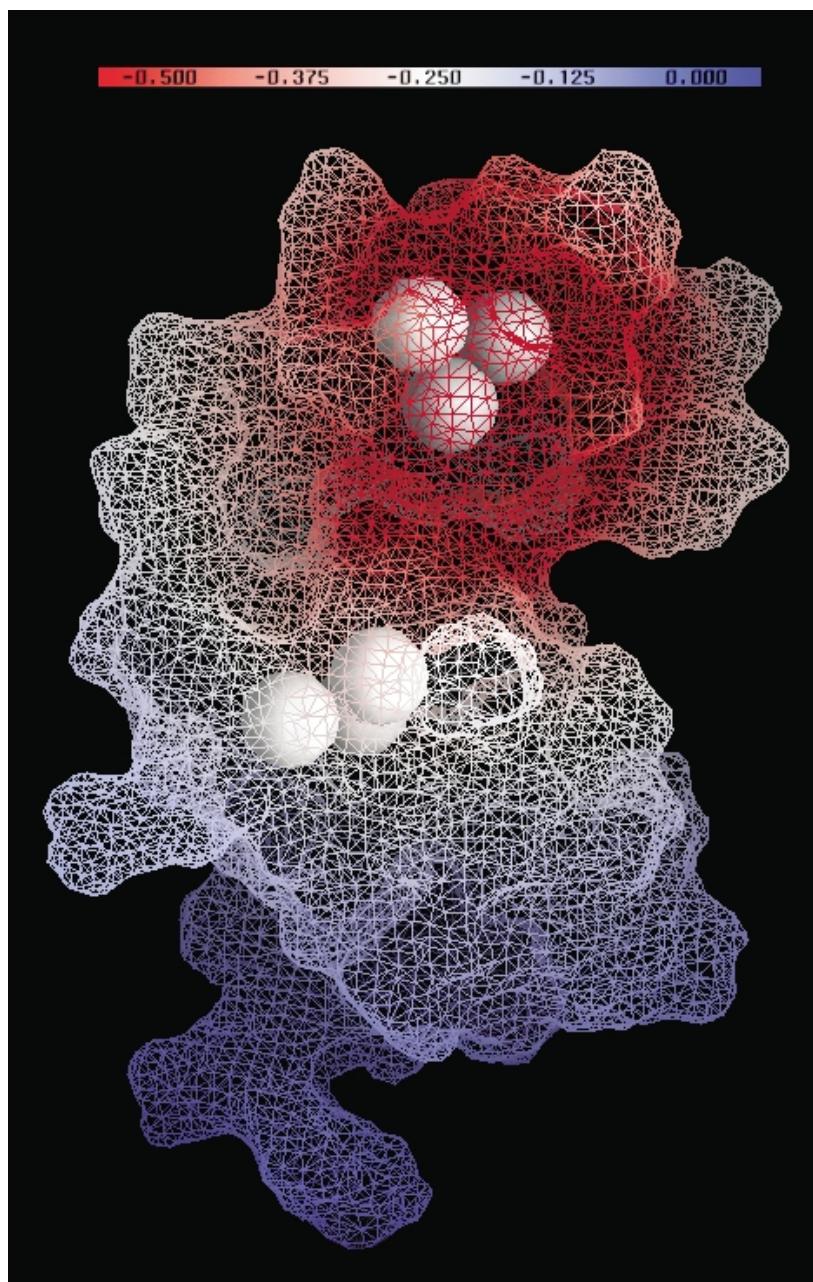


Figure 4. The local hardness map of NC. The surface is rotated to give the best view of the solvent-exposed Zn^{2+} -coordinating sulfur atoms (shown as white spheres). The color scheme is -0.5 (red) to 0.0 (blue). The local hardness is given in units of kT/e ($T = 298$ K).

the unbound and bound form are small, of the order of $10^{-2} e$ (e is atomic unit of charge). In both the unbound and the RNA-bound complex, the Cys15 sulfur atom of ZF1 carries the most negative charge, and the Cys49 sulfur of ZF2 the least negative charge. Variations in charge reflect, in part, difference in the electrostatic potential in the polarized macromolecule. Negative charge will tend to accumulate in more electropositive regions and *vice versa*. The atomic charge variations observed here are consistent with a recent Poisson–Boltzmann study that uses a static charge model,⁴³ which found that the Cys49 thiolate experienced the least positive screening from the protein

environment, whereas Cys36 and Cys15 thiolate groups experienced the most positive screening (also apparent from the electrostatic potential map in Figure 1). The relative magnitude of the zinc finger thiolate charges is a result of the interplay of the solvent polarization with the protein electrostatic environment. Despite the fact that the Cys49 thiol is solvent-exposed, it resides in close proximity to the negatively charged neighbor Asp48. The net effect of this is to destabilize the negative charge on the thiolate relative to that on the other zinc-binding thiolate groups.

RNA binding results in a small change (less than $0.05 e$) in the thiolate partial charges. The largest

Table 1. Atomic charges on the Zn²⁺-coordinating atoms in NC and the NC-SL3 RNA complex

Residue	Atom	Charge (NC)	Charge (NC-SL3)
Cys15	S ^γ	-0.512(5.63 × 10 ⁻³)	-0.489(6.67 × 10 ⁻³)
Cys18	S ^γ	-0.459(2.24 × 10 ⁻³)	-0.469(7.74 × 10 ⁻³)
His23	N ^{ε2}	0.062(8.76 × 10 ⁻³)	0.066(9.99 × 10 ⁻³)
Cys28	S ^γ	-0.415(7.07 × 10 ⁻³)	-0.443(6.79 × 10 ⁻³)
Cys36	S ^γ	-0.478(4.94 × 10 ⁻³)	-0.432(4.86 × 10 ⁻³)
Cys39	S ^γ	-0.476(8.14 × 10 ⁻³)	-0.486(9.19 × 10 ⁻³)
His44	N ^{ε2}	0.065(6.10 × 10 ⁻³)	0.067(6.86 × 10 ⁻³)
Cys49	S ^γ	-0.408(1.39 × 10 ⁻²)	-0.424(1.45 × 10 ⁻²)

CM2 charges⁷⁶ derived from linear-scaling calculations in solution using PM3 Hamiltonian.⁷⁸ RMS deviations due to conformational variations in the NMR ensemble structure are included in parentheses. Units are in *e*.

change occurs on Cys36, increasing the thiolate charge from -0.478 *e* in NC to -0.432 *e* in the NC-SL3 RNA complex. For both unbound and SL3 RNA-bound protein, the RMS deviations of partial charges for the conformational variations in the NMR ensemble structures are of the order of 10⁻³ *e*, with the exception of the Cys49 thiolate group, where the RMSD is about 0.01 *e*. The noticeable charge variation on the Cys49 thiolate group is consistent with the pronounced conformational flexibility of the nearby carboxyl terminus,²⁰ and the ability of the chemically softer Cys49 thiolate to accommodate local charge fluctuations,⁵⁰ as can be seen in the discussions on chemical reactivity indices.

Chemical reactivity indices

The local hardness, $\eta(\vec{r})$, is a measure of chemical reactivity at a position \vec{r} , and satisfies a reciprocal relation with the local softness (see Methods). In this work, the local hardness is approximated as the electrostatic potential due to a charge density related to the local softness^{51,52} (i.e. not of the molecular charge density). The local hardness and atomic softness indices are a set of properties of

Table 2. Atomic softness indices on the Zn²⁺-coordinating atoms in NC and the NC-SL3 RNA complex

Residue	Atom	NC	NC-SL3
Cys15	S ^γ	1.504(0.998)	0.079(0.038)
Cys18	S ^γ	1.758(1.265)	0.023(0.017)
His23	N ^{ε2}	0.022(0.005)	0.017(0.002)
Cys28	S ^γ	0.493(0.410)	0.018(0.004)
Cys36	S ^γ	1.657(1.022)	0.479(1.290)
Cys39	S ^γ	0.972(1.097)	0.245(0.729)
His44	N ^{ε2}	0.060(0.025)	0.024(0.026)
Cys49	S ^γ	5.101(2.139)	1.496(1.800)

Atomic softness indices (J.K. & D.M.Y., unpublished results) derived from linear-scaling calculations in solution using the PM3 Hamiltonian.⁷⁸ RMS deviations due to conformational variations in the NMR ensemble structure are included in parentheses. Units are in *e*.

similar type to the molecular electrostatic potential and atomic charges, but they provide information about local chemical reactivity of the molecule. In an analogous way to the electrostatic potential map described earlier, a “local hardness map” on the solvent-accessible (or solvent-excluded) molecular surface can be visualized graphically and used to identify regions of greater nucleophilicity (i.e. large magnitude of the local hardness). The local hardness map of NC (Figure 4) shows that the surface region around ZF2 displays a larger local hardness value (indicated in red) than the ZF1 region. This indicates that the electron density of the highest-lying occupied molecular orbitals (HOMOs) are more concentrated in ZF2, and consequently that ZF2 is likely to be more susceptible to electrophilic attack than ZF1.

The atomic local softness indices provide a convenient measure of nucleophilicity of individual atoms, with the atoms exhibiting a large value of local softness index expected to be more reactive. Table 2 shows the average and RMS deviations of the atomic local softness indices on the Zn²⁺-coordinating atoms of NC and NC-SL3 RNA. The largest average atomic local softness index resides on the Cys49 thiolate (0.227), and is significantly larger than those on other thiolate groups, indicating the Cys49 thiolate to be the most nucleophilic among all the zinc-binding thiolate groups. Additional calculations (data not shown) showed that the HOMO indeed originates from Cys49, with the largest orbital coefficient located on the sulfur atom of this residue. The large RMS deviation of the local softness index on the Cys49 thiolate is consistent with the pronounced charge fluctuation on this atom as discussed earlier. A large conformation-dependent fluctuation of the local softness on Cys49 thiolate was also observed in a gas-phase density-functional study⁴² based on isolated model zinc finger structures derived from an NMR structure of the free NC (PDB 1AAF). Evidently, the conformational fluctuation of the local protein environment can have a profound impact on the frontier orbitals. In the RNA-bound protein, as expected, the magnitude of local softness indices on the Zn²⁺-coordinating atoms is much smaller (Table 2), since the high-lying occupied orbitals are located mainly on the negatively charged RNA. Nevertheless, within the protein, the set of most reactive atoms remains the same.

The local hardness map and local softness indices both suggest that ZF2, and in particular the Cys49 sulfur atom, is the most susceptible toward attack by electrophilic agents. These results are in accord with the experimental observation that Cys49 is the most reactive cysteine of NC toward small electrophilic drug candidates^{37,38} and are consistent with the previous study based on density-functional theory.⁴² The results presented here underscore the potential applications of new information derivable from linear-scaling electronic structure methods in structure-based drug design.

Table 3. Comparison of calculated relative proton potentials and experimental pK_a values of Zn²⁺-coordinating residues in the NC apoprotein

Residue	pK _a ^a	ϕ
Cys36	8.0 ± 0.2	-3.278
Cys39	8.8 ± 0.2, 8.2 ± 0.3, 8.7 ± 0.3	-3.241
His44	6.4 ± 0.1	-0.217
Cys49	9.3 ± 0.2	-5.053

ϕ is the relative electrostatic potential (V) at the proton from thiol or imidazole of the Zn²⁺-coordinating residues in the NC apoprotein.

^a Experimental values were obtained from ¹H NMR spectroscopic studies of the (35-50)NC apopeptide.⁵⁶

Relative proton potentials

The electrostatic potential at a proton position (neglecting the contribution of the proton itself) is a useful index that reveals the relative stability of the proton. Consequently, for macromolecules, the potentials at the titratable proton positions offer information about the relative acidity and basicity for a particular protonation state. In systems where titratable sites are weakly coupled, proton potentials may provide a useful indication of the ordering of pK_a values. (Note, this is not the same as the calculation of pK_a shifts or a complete titration curve that require an ensemble average over all relevant protonation states, a typically much larger calculation.^{53,54}) Table 3 gives the computed relative proton potentials and experimental pK_a values of the zinc-binding thiol groups in ZF2 of the NC apoprotein. These quantities appear to be well correlated, although the correlation may not be statistically significant. Notably, the proton on the thiol of Cys49 has the most negative potential and the one on the imidazole of His44 the least. The trend in the relative magnitudes of proton potentials correlates well with the order of measured pK_a values on the Zn²⁺-coordinating residues of the (35-50)NC peptide, a model peptide corresponding to ZF2.⁵⁵ The significantly more negative proton potential on the Cys49 thiol can be attributed to the negative charge of the neighboring residue Asp48 that has been suggested elsewhere⁵⁵ to give rise to the unusually high pK_a of Cys49. The theoretical results in the present work support the model suggested by Bombarda *et al.*, that Cys49 may act as a switch in a stepwise Zn²⁺ dissociation mechanism of ZF2^{55,56} and offers yet another rationale for the high susceptibility of Cys49 to electrophilic attack.

Protein side-chain contributions to the electrostatic-binding energy

Finally, we examine the distinctive binding mode of NC to the SL2 and SL3 RNA stem-loops of the packaging signal on the basis of protein side-chain contributions to the electrostatic binding free energies. The protein backbone contributions

are relatively small (data not shown) in comparison to those from the side-chains and will not be considered here. This study was carried out using the two available high-resolution NMR structures of NC-SL2 RNA²¹ and NC-SL3 RNA.²⁰ Protein side-chain contribution to the electrostatic binding energy is defined as the change in the electrostatic component of the binding free energy (including the solvent contribution) upon switching off the charge on the corresponding side-chain group (isosteric group). Thus, a positive contribution implies a stabilization of the NC–RNA complex. Here, the approximation is made such that the structure of the protein and RNA in the unbound (dissociated) state remains unchanged from their bound state. Hence, the present method does not take into account the conformational relaxation that occurs upon dissociation. Nevertheless, the “rigid structure” approach has been proved useful in estimating the electrostatic contribution of salt-bridges to protein stability,^{57–59} protein–protein binding^{60,61} and ligand–DNA interactions.⁶² Considering that the NC–RNA-binding free energy is highly dependent on salt concentration,^{18,63–65} and the binding is mainly electrostatically driven at low ionic strength,⁶⁶ it is apparent that electrostatics play an essential role in NC–RNA binding. Thus, examination of the electrostatic component of the binding free energy in terms of individual amino acid side-chain contributions can provide insight into the nature of specific NC–RNA interactions and the origin of conservation of certain amino acid residues.

Figure 5 shows the average and conformational root-mean-squared deviations of amino acid side-chain contributions to the electrostatic binding free energy of NC-SL2 (upper) and NC-SL3 (lower). While the NC-SL3 binding is mainly stabilized through the interaction of the 3₁₀ α-helical side-chains with the RNA, the NC SL2-RNA

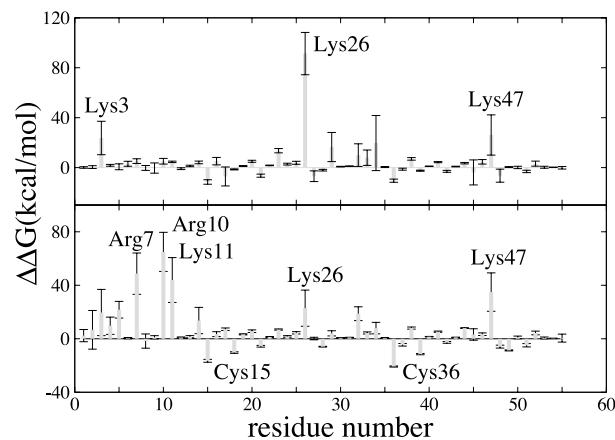


Figure 5. NC amino acid side-chain contributions to the electrostatic-binding energy in NC:SL2 RNA (top) and NC:SL3 RNA (bottom). The error bars indicate the RMS deviations in $\Delta\Delta G^{\text{el}}$ due to conformational variations in the NMR ensemble structure. The labeled residues are discussed in the text.

binding seems to be dominated by the contribution from the side-chain of Lys26, which forms a salt-bridge with the negatively charged RNA phosphodiester backbone. The large electrostatic stabilization from the conserved basic residues, Arg7, Arg10, and Lys11 of the 3_{10} α -helix in the NC-SL3 complex, is a consequence of the favorable electrostatic contacts between their side-chains and the RNA phosphodiester backbone, which are absent from the SL2 complex as they project away from the RNA. However, electrostatic contact is not the only determining factor, as solvent and local electrostatic environment can also modulate the electrostatic interaction. The Lys26 side-chain lies on the interface between the SL2-RNA and NC protein, whereas it is partially exposed to solution in the SL3 complex. Not surprisingly, the most destabilizing groups in both complexes are the Cys15 and Cys36 side-chains, the sulfur atoms of which bear the most negative charges according to our calculations (Table 1).

The role of basic residues in NC functions has been investigated experimentally. Mutageneses of several basic residues in the N-terminal region including Arg7, Arg10, and Lys11 resulted in a strong decrease in viral infectivity with an observed defect in proviral DNA synthesis.⁶⁷ *In vitro* studies have also revealed an important role for NCs basic residues in nucleic acid binding.^{66,68} An early study showed Arg residues at positions 3, 7, 10, 29, and 32 are essential for the specific binding of NC to the viral RNA-packaging site.⁶⁸ Based on fluorescent studies with poly(eA), NC residues Arg7, Lys20, Lys26, and Lys41 were proposed to be involved in electrostatic contacts with the RNA.⁶⁶ Our data generally support these experimental findings. Although sizable RMS deviations in the calculated values due to the conformational fluctuations in the NMR ensemble are observed, the deviations do not change the relative contribution of individual residues to the binding energy.

Conclusion

Here, the regioselective reactivity of HIV-1 NC toward electrophilic attack is investigated with a set of quantum mechanically derived macromolecular descriptors. The electrostatic potential map reveals the ZF2 region to be more electronegative than the ZF1 region, making it more accessible by electrophiles. Upon RNA binding, the potential in the ZF2 region becomes less electronegative and resembles that in the ZF1 region. Among all six Zn^{2+} -coordinating sulfur atoms in the unbound protein, the sulfur atom of Cys49 carries the least negative charge, and the sulfur of Cys15 the greatest. The destabilization of the negative charge on Cys49 can be mainly attributed to the negatively charged neighboring residue Asp48. The surface region around ZF2 exhibits an

increased value of local hardness, indicating it to be the most reactive region toward electrophiles.

The largest value of the local softness index is found on the Cys49 sulfur atom, offering a rationale for its particularly high susceptibility toward electrophilic attack, consistent with experimental observations.^{37,38} Large softness indices are also found on other Zn^{2+} -coordinating thiolate groups, suggesting their increased reactivity as compared to the rest of the atoms in the protein. Upon RNA binding, the magnitude of local softness indices on the Zn^{2+} -coordinating thiolate groups decreases but they remain the most nucleophilic atoms within the protein.

The relative proton potentials of the ZF2 Zn^{2+} -coordinating residues in the apoprotein have been shown to correlate well with the experimental pK_a values. In particular, the most negative relative proton potential on Cys49 is consistent with its unusually high pK_a value, which again can be attributed to the acidic neighbor Asp48, and supports the hypothesis that Cys49 may act as a switch for a stepwise Zn^{2+} dissociation mechanism in an acidic cellular environment.^{56,57}

The electrostatic binding energy calculations emphasize the importance of several basic residues in the 3_{10} helical region and ZF1 of NC for stabilizing binding to SL3 RNA, whereas Lys26 appears to be the most important residue in the electrostatic binding of NC to SL3 RNA. The distinctive electrostatic binding mode of NC to SL2 and SL3-RNA revealed from our calculations is striking and support the idea^{20,21} that NC–RNA binding is adaptive and occurs *via* different sets of interactions, in accord with NC's multiple nucleic acid-binding functions in the viral life-cycle. These results also help to rationalize the presence of evolutionarily conserved residues in NC. The macromolecular descriptors and electrostatic binding energy calculations reported here provide new insights that may be useful in the design and optimization of anti-HIV therapeutics.

Methods

Structure preparation

The NMR ensemble of 25 NC:SL3–RNA complexes (PDB 1A1T, about 1500 atoms including hydrogen atoms) were used as starting structures. The hydrogen positions of the protein:RNA complexes were optimized using the CHARMM27 program^{69,70} with the CHARMM22 force-field for proteins,⁷¹ and CHARMM27 force-field for nucleic acids.⁷² The parameters for zinc were taken from Stote & Karplus.⁷³ The non-bond interactions were evaluated with a distance-dependent dielectric function ($\epsilon = r$) and no cut-offs (no explicit water molecules were included). Each structure was relaxed with 100 steps of steepest descent on a semiempirical quantum surface using the linear-scaling divide-and-conquer semiempirical program with the linear-scaling COSMO solvation method.^{44,74} The PM3 Hamiltonian was used as the PM3 parameterization was previously

shown to give experimentally agreeable geometries for zinc complexes.⁷⁵ In the linear-scaling quantum geometry optimization, the subsystems were chosen to be single amino acid or nucleotide residues and the buffer/matrix cut-off of 6 Å/7 Å were applied. The SCF convergence criterion was set at 10⁻⁵ kcal mol⁻¹ atom⁻¹. The same procedure was applied to obtain the quantum optimized structures from the NMR ensemble of 20 NC:SL2–RNA complexes (PDB 1F6U). The relaxed structures were used in the binding energy calculations.

Atomic softness index and local hardness map

Here, the calculations of the local hardness and the atomic local softness indices are briefly described. A detailed discussion of the methods, implementation and test-case validation will be given elsewhere. Chemical hardness and softness are two fundamental quantities in inorganic chemistry.⁵⁰

The chemical hardness η is defined rigorously in density-functional theory as the second derivative of the energy with respect to the total number of electrons with the external nuclear potential v held fixed (in some conventions, a factor of 1/2 is included). A commonly employed measure of the chemical hardness, which can be derived with a finite difference approximation is the difference between the ionization potential (IP) and electron affinity (EA):^{51,52}

$$\eta = \left[\frac{\partial^2 E}{\partial N^2} \right]_v \approx IP - EA \quad (1)$$

The chemical softness is the inverse of the hardness $S = 1/\eta$.

A related quantity defined in density-functional theory that has a close connection with frontier orbital theory is the Fukui function, defined as $f(\vec{r}) = [\partial\rho(\vec{r})/\partial N]_v$. The Fukui function describes the sensitivity of the electron density with respect to total number of electrons, and integrates to unity ($\int f(\vec{r})d^3r = 1$). Sometimes the left $f^-(\vec{r})$ and right $f^+(\vec{r})$ derivatives are distinguished to emphasize that electrons being taken from the system (left derivative) may come from a different region than where those added to the system (right derivative) may go; the middle derivative defined as the average $f^0(\vec{r}) = 1/2[f^+(\vec{r}) + f^-(\vec{r})]$. In other words, if electrons are added to (or taken from) the system, the Fukui function $f^+(\vec{r})$ (or $f^-(\vec{r})$) indicates spatially where the electron density is likely to increase (or decrease) as a result. Consequently, the Fukui function can be regarded as a reactivity index that provides information about the spatial susceptibility of nucleophilic (f^+), electrophilic (f^-) or radical (f^0) attack. The Fukui function can be approximated using a finite difference/frozen orbital approximation as:

$$f(\vec{r})^\pm \approx \pm \frac{1}{\Delta N} [\rho(\vec{r}; N \pm \Delta N) - \rho(\vec{r}; N)] \quad (2)$$

where $f(\vec{r})^-$, $f(\vec{r})^+$ are the left and right Fukui functions that give rise to probable sites for electrophilic and nucleophilic attack, respectively. From a frontier-orbital perspective, the left and right Fukui functions represent approximately the HOMO and LUMO densities, respectively. For biological macromolecules, due to a large number of electrons and many nearly degenerate states close to the frontier orbitals, a ΔN value larger than 1 is used in the finite difference approximation ($\Delta N = 5$ was used in the current work).

The local softness is related to the chemical softness S

and Fukui function(s) $f(\vec{r})$ by $s(\vec{r}) = Sf(\vec{r})$. By applying an electron density partition scheme, such as Mulliken population or CM2 charge mapping⁷⁶ (used in this work), the Fukui function and local softness can be “assigned” to atoms resulting in an atomic Fukui index and atomic softness index, respectively. The present work focuses on electrophilic attack of the NC zinc fingers, and hence the local softness index derived from f^- are used throughout the discussion.

A quantity related to the local softness $s(\vec{r})$ is the local hardness $\eta(\vec{r})$. The local hardness can be approximated as the electrostatic potential due to the Fukui function:

$$\eta(\vec{r}) \approx \int \frac{f(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' \quad (3)$$

For visualization purpose, it can be mapped onto a molecular surface in the same fashion as the molecular electrostatic potential. A surface area with a large magnitude of local hardness implies a region of increased chemical reactivity.

Calculation of electrostatic properties, reactivity indices and electrostatic binding energies

The 19 structures with the lowest heat of formation (less than -10,000 kcal/mol) after 100 steps of steepest descent quantum energy minimizations were chosen for the subsequent single-point linear-scaling PM3 calculations of the atomic charges, electrostatic potential, atomic softness indices and local hardness on the isolated NC protein and the SL3-RNA bound NC protein. Here, a buffer/matrix cut-off of 8 Å/9 Å was used. In the atomic charge and atomic softness index calculations, the CM2 charge model⁷⁶ was used. The solvent-polarized potentials were computed on a 64 × 64 × 64 grid and then mapped onto a molecular surface and graphically displayed with program GRASP.⁷⁷

In the NC:RNA-binding energy calculations, the linear-scaling smooth COSMO program was used.^{45,74} The electrostatic-binding free energy was computed as:

$$\Delta G_{\text{bind}}^{\text{el}} = G_{\text{tot}}^{\text{el}}(\text{protein:RNA}) - [G_{\text{tot}}^{\text{el}}(\text{protein}) + G_{\text{tot}}^{\text{el}}(\text{RNA})] \quad (4)$$

where $G_{\text{tot}}^{\text{el}} = G_{\text{gas}}^{\text{el}} + \Delta G_{\text{solvation}}^{\text{el}}$, $G_{\text{gas}}^{\text{el}}$ is the gas-phase electrostatic energy, and $\Delta G_{\text{solvation}}^{\text{el}}$ is the electrostatic component of the solvation free energy. The protein side-chain contribution to the electrostatic binding energy for the i th amino acid were obtained as $\Delta G_i^{\text{el}} = \Delta G(\text{off})_i - \Delta G(\text{on})_i$, where $\Delta G(\text{off})_i$ and $\Delta G(\text{on})_i$ represent ΔG_{bind} with the charges on the i th amino acid side-chain “switched” off and on, respectively.

Acknowledgements

We acknowledge financial support from the National Institutes of Health (grant GM62248 to D.Y. and AI65056 to K.M.F.) and the Donors of The Petroleum Research Fund, administered by the American Chemical Society. J.K. was partially supported by the Louise T. Dosdall Fellowship from the University of Minnesota. Computational resources were provided by the Minnesota Supercomputing Institute.

References

- Tsuchihashi, Z. & Brown, P. O. (1994). DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **68**, 5863–5870.
- Rein, A., Henderson, L. E. & Levin, J. G. (1998). Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* **23**, 297–301.
- Darlix, J.-L., Lapadat-Tapolsky, M., de Rocquigny, H. & Roques, P. B. (1995). First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–537.
- Herschlag, D. (1995). RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **270**, 20871–20874.
- Lorsch, J. R. (2002). RNA chaperones exist and DEAD box proteins get a life. *Cell*, **109**, 797–801.
- de Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M.-C., Roques, B. & Darlix, J.-L. (1992). Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl Acad. Sci. USA*, **89**, 6472–6476.
- Tisné, C., Roques, B. P. & Dardel, F. (2001). Heteronuclear NMR studies of the interaction of tRNA^{Lys,3} with HIV-1 nucleocapsid protein. *J. Mol. Biol.* **306**, 443–454.
- Hargittai, M. R. S., Mangla, A. T., Gorelick, R. J. & Musier-Forsyth, K. (2001). HIV-1 nucleocapsid protein zinc finger structures induce tRNA^{Lys,3} structural changes but are not critical for primer/template annealing. *J. Mol. Biol.* **312**, 985–997.
- Allain, B., Lapadat-Tapolsky, M., Berlioz, C. & Darlix, J.-L. (1994). Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome. *EMBO J.* **13**, 973–981.
- You, J. C. & McHenry, C. S. (1994). Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **269**, 31491–31495.
- Rodríguez-Rodríguez, L., Tsuchihashi, Z., Fuentes, G. M., Bambara, R. A. & Fay, P. J. (1995). Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase *in vitro*. *J. Biol. Chem.* **270**, 15005–15011.
- Guo, J., Henderson, L. E., Bess, J., Kane, B. & Levin, J. G. (1997). Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* **71**, 5178–5188.
- Wu, T., Guo, J., Bess, J., Henderson, L. E. & Levin, J. G. (1999). Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. *J. Virol.* **73**, 4794–4805.
- Auxilien, S., Keith, G., Le Grice, S. F. J. & Darlix, J.-L. (1999). Role of post-transcriptional modifications of primer tRNA^{Lys,3} in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. *J. Biol. Chem.* **274**, 4412–4420.
- Berg, J. M. (1986). Potential metal-binding domains in nucleic acid binding proteins. *Science*, **232**, 485–487.
- Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J. et al. (2000). Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* **74**, 8980–8988.
- Guo, J., Wu, T., Kane, B. F., Johnson, D. G., Henderson, L. E., Gorelick, R. J. & Levin, J. G. (2002). Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **76**, 4370–4378.
- Williams, M. C., Rouzina, I., Wenner, J. R., Gorelick, R. J., Musier-Forsyth, K. & Bloomfield, V. A. (2001). Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl Acad. Sci. USA*, **98**, 6121–6126.
- Williams, M. C., Gorelick, R. J. & Musier-Forsyth, K. (2002). Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl Acad. Sci. USA*, **99**, 8614–8619.
- De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science*, **279**, 384–388.
- Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R. & Summers, M. F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **301**, 491–511.
- Gorelick, R. J., Nigida, S. M., Jr, Bess, J. W., Jr, Arthur, L. O., Henderson, L. E. & Rein, A. (1990). Non-infectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–3211.
- Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E. & Arthur, L. O. (1993). The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J. Virol.* **67**, 4027–4036.
- Gorelick, R. J., Fu, W., Gagliardi, T. D., Bosche, W. J., Rein, A., Henderson, L. E. & Arthur, L. O. (1999). Characterization of the block in replication of nucleocapsid protein zinc finger mutants from oloney murine leukemia virus. *J. Virol.* **73**, 8185–8195.
- Buckman, J. S., Bosche, W. J. & Gorelick, R. J. (2003). Human immunodeficiency virus type 1 nucleocapsid Zn²⁺-fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J. Virol.* **77**, 1469–1480.
- Déméné, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N. et al. (1994). ¹H NMR structure and biological studies of the His²³-Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry*, **33**, 11707–11716.
- Tanchou, V., Decimo, D., Péchoux, C., Lener, D., Rogemond, V., Berthoux, L., Ottmann, M. & Darlix, J.-L. (1998). Role of the N-terminal zinc finger of human immunodeficiency virus type 1 nucleocapsid protein in virus structure and replication. *J. Virol.* **72**, 4442–4447.
- Rice, W. G., Schaeffer, C. A., Harten, B., Villinger, F., South, T. L., Summers, M. F. et al. (1993). Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds. *Nature*, **361**, 473–475.

29. Tummino, P. J., Scholten, J. D., Harvey, P. J., Holler, T. P., Maloney, L., Gogliotti, R. *et al.* (1996). The *in vitro* ejection of zinc from human immunodeficiency virus (HIV) type 1 nucleocapsid protein by disulfide benzamides with cellular anti-HIV activity. *Proc. Natl Acad. Sci. USA*, **93**, 969–973.
30. Huang, M., Maynard, A., Turpin, J. A., Graham, L., Janini, G. M., Covell, D. G. & Rice, W. G. (1998). Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **41**, 1371–1381.
31. Turpin, J. A., Song, Y., Inman, J. K., Huang, M., Wallqvist, A., Maynard, A. *et al.* (1999). Synthesis and biological properties of novel pyridinioalkanoyl thioesters (PATE) as anti-HIV-1 agents that target the viral nucleocapsid protein zinc fingers. *J. Med. Chem.* **42**, 67–86.
32. Druillennec, S., Dong, C. Z., Escaich, S., Gresh, N., Bousseau, A., Roques, B. P. & Fournié-Zaluski, M. C. (1999). A mimic of HIV-1 nucleocapsid protein impairs reverse transcription and displays antiviral activity. *Proc. Natl Acad. Sci. USA*, **96**, 4886–4891.
33. Stephen, A. G., Worthy, K. M., Towler, E., Mikovits, J. A., Sei, S., Roberts, P. *et al.* (2002). Identification of HIV-1 nucleocapsid protein:nucleic acid antagonists with cellular anti-HIV activity. *Biochem. Biophys. Res. Commun.* **296**, 1228–1237.
34. Rice, W. G., Schaeffer, C. A., Graham, L., Bu, M., McDougal, J. S., Orloff, S. L. *et al.* (1993). The site of antiviral action of 3-nitrosobenzamide on the infectivity process of human immunodeficiency virus in human lymphocytes. *Proc. Natl Acad. Sci. USA*, **90**, 9721–9724.
35. Rice, W. G., Supko, J. G., Malspeis, L., Buckheit, R. W., Jr, Clanton, D., Bu, M. *et al.* (1995). Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of aids. *Science*, **270**, 1194–1197.
36. Rice, W. G., Turpin, J. A., Huang, M., Clanton, D., Buckheit, R. W., Jr, Covell, D. G. *et al.* (1997). Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nature Med.* **3**, 341–345.
37. Hathout, Y., Fabris, D., Han, M. S., Sowder, R., II, Henderson, L. E. & Fenselau, C. (1997). Characterization of intermediates in the oxidation of zinc fingers in human immunodeficiency virus type 1 nucleocapsid protein P7. *Drug Metab. Dispos.* **24**, 1395–1400.
38. Chertova, E. N., Kane, B. P., McGrath, C., Johnson, D. G., Sowder, R. C., II, Arthur, L. O. & Henderson, L. E. (1998). Probing the topography of HIV-1 nucleocapsid protein with the alkylating agent *N*-ethyl-maleimide. *Biochemistry*, **37**, 17890–17897.
39. Basrur, V., Song, Y., Mazur, S. J., Higashimoto, Y., Turpin, J. A., Rice, W. G. *et al.* (2000). Inactivation of HIV-1 nucleocapsid protein p7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action. *J. Biol. Chem.* **275**, 14890–14897.
40. Topol, I. A., Nemukhin, A. V., Chao, M., Iyer, L. K., Tawa, G. J. & Burt, S. K. (2000). Quantum chemical studies of reactions of the cyclic disulfides with the zinc finger domains in the HIV-1 nucleocapsid protein (NCp7). *J. Am. Chem. Soc.* **122**, 7087–7094.
41. Topol, I. A., Nemukhin, A. V., Dobrogorskaya, Y. I. & Burt, S. K. (2001). Interactions of azodicarbonamide (ADA) species with the model zinc finger site: theoretical support of the zinc finger domain destruction in the HIV-1 nucleocapsid protein (NCp7) by ADA. *J. Phys. Chem. ser. B*, **105**, 11341–11350.
42. Maynard, A. T., Huang, M., Rice, W. G. & Covell, D. G. (1998). Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc. Natl Acad. Sci. USA*, **95**, 11578–11583.
43. Maynard, A. T. & Covell, D. G. (2001). Reactivity of zinc finger cores: analysis of protein packing and electrostatic screening. *J. Am. Chem. Soc.* **123**, 1047–1058.
44. York, D. M., Lee, T.-S. & Yang, W. (1996). Parameterization and efficient implementation of a solvent model for linear-scaling semiempirical quantum mechanical calculations of biological macromolecules. *Chem. Phys. Letters*, **263**, 297–306.
45. York, D. M. & Karplus, M. (1999). A smooth solvation potential based on the Conductor-Like Screening model. *J. Phys. Chem. ser. A*, **103**, 11060–11079.
46. Goedecker, S. (1999). Linear scaling electronic structure methods. *Rev. Mod. Phys.* **71**, 1085–1123.
47. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
48. Gilson, M. K. (1995). Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **5**, 216–223.
49. Chin, K., Sharp, K. A., Honig, B. & Pyle, A. M. (1999). Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. *Nature Struct. Biol.* **6**, 1055–1061.
50. Pearson, R. G. (1997). *Chemical Hardness*, Wiley-VCH, Weinheim.
51. Parr, R. & Yang, W. (1989). *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York.
52. Chermette, H. (1999). Chemical reactivity indexes in density functional theory. *J. Comput. Chem.* **20**, 129–154.
53. Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. (1993). On the calculation of pK_a 's in proteins. *Proteins: Struct. Funct. Genet.* **15**, 252–265.
54. Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1996). The determinants of pK_a 's in proteins. *Biochemistry*, **35**, 7819–7833.
55. Bombarda, E., Morellet, N., Cherradi, H., Spiess, B., Bouaziz, S., Grell, E. *et al.* (2001). Determination of the pK_a of the four Zn^{2+} -coordinating residues of the distal finger motif of the HIV-1 nucleocapsid protein: consequences on the binding of Zn^{2+} . *J. Mol. Biol.* **310**, 659–672.
56. Bombarda, E., Cherradi, H., Morellet, N., Roques, B. P. & Mély, Y. (2002). Zn^{2+} -binding properties of single-point mutants of the C-terminal zinc finger of the HIV-1 nucleocapsid protein: evidence of a critical role of Cysteine 49 in Zn^{2+} dissociation. *Biochemistry*, **41**, 4312–4320.
57. Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226.
58. Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435–1444.
59. Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**, 1241–1255.
60. Sheinerman, F. B., Norel, R. & Honig, B. (2000). Electrostatic aspects of protein protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159.
61. Norel, R., Sheinerman, F., Petrey, D. & Honig, B. (2001). Electrostatic contributions to protein protein

- interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* **10**, 2147–2161.
62. Misra, V. K. & Honig, B. (1995). On the magnitude of the electrostatic contribution to ligand–DNA interactions. *Proc. Natl Acad. Sci. USA*, **92**, 4691–4695.
 63. Lapadat-Tapolsky, M., Pernelle, C., Borie, C. & Darlix, J. L. (1995). Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucl. Acids Res.* **23**, 2434–2441.
 64. Mély, Y., de Rocquigny, H., Sorinas-Jimeno, M., Keith, G., Roques, B. P. & Marquet, R. (1995). Binding of the HIV-1 nucleocapsid protein to the primer tRNA, *in vitro*, is essentially not specific. *J. Biol. Chem.* **270**, 1650–1656.
 65. Vuilleumier, C., Bombarda, E., Morellet, N., Gérard, D., Roques, B. P. & Mély, Y. (1999). Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry*, **38**, 16816–16825.
 66. Urbaneja, M. A., Kane, B. P., Johnson, D. G., Gorelick, R. J., Henderson, L. E. & Casas-Finet, J. R. (1999). Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *J. Mol. Biol.* **287**, 59–75.
 67. Berthoux, L., Péchoux, C., Ottmann, M., Morel, G. & Darlix, J.-L. (1997). Mutations in the N-terminal domain of human immunodeficiency virus type 1 nucleocapsid protein affect virion core structure and proviral DNA synthesis. *J. Virol.* **71**, 6973–6981.
 68. Dannull, J., Surovoy, A., Jung, G. & Moelling, K. (1994). Specific binding of HIV-1 nucleocapsid protein to Ψ RNA *in vitro* requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.* **13**, 1525–1533.
 69. Brooks, B. R., Brucolieri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). Charmm: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
 70. MacKerell, A. D. Jr, Brooks, B., Brooks, C. L. III, Nilsson, L., Roux, B., Won, Y. & Karplus, M. (1998). CHARMM: the energy function and its parameterization with an overview of the program. In *Encyclopedia of Computational Chemistry* (Schleyer, P., Allinger, N., Clark, T., Gasteiger, J., Kollman, P., Schaefer, H. III & Schreiner, P., eds), vol. 1, pp. 271–277, Wiley, Chichester.
 71. MacKerell, A. D., Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J. D., Field, M. J. et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. ser. B*, **102**, 3586–3616.
 72. Foloppe, N. & MacKerell, A. D., Jr (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104.
 73. Stote, R. H. & Karplus, M. (1995). Zinc binding in proteins and solution: a simple but accurate non-bonded representation. *Proteins: Struct. Funct. Genet.* **23**, 12–31.
 74. Klamt, A. & Schüürmann, G. (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, **2**, 799–805.
 75. Lewis, J. P., Liu, S., Lee, T.-S. & Yang, W. (1999). A linear-scaling quantum mechanical investigation of cytidine deaminase. *J. Comput. Phys.* **151**, 242–263.
 76. Li, J., Zhu, T., Cramer, C. J. & Truhlar, D. G. (1998). New class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. ser. A*, **102**, 1820–1831.
 77. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.*, **11**, 281–296.
 78. Stewart, J. J. P. (1989). Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **10**, 209–220.

Edited by M. F. Summers

(Received 6 March 2003; received in revised form 15 May 2003; accepted 16 May 2003)