Reactivity of Zinc Finger Cores: Analysis of Protein Packing and Electrostatic Screening

A. T. Maynard*,† and D. G. Covell*

Contribution from the Frederick Cancer Research and Development Center, Screening Technologies Branch, DCTD, NCI, NIH, SAIC, Frederick, Maryland 21702

Received April 3, 2000. Revised Manuscript Received October 23, 2000

Abstract: The chemical stability of 207 zinc fingers, derived from 92 experimental protein structures, is evaluated according to the protein packing and electrostatic screening of their zinc cores. These properties are used as measures of the protein protection of zinc cores, to predictively rank relative zinc finger reactivities and assess differences in function. On average, there is a substantial and concomitant increase in the screening of increasingly anionic core motifs, suggesting zinc fingers have evolved in a manner that promotes shielding of their potentially reactive core thiolates. In contrast, *enzymatic* zinc cores are functionally differentiated by negative electrostatic screening. Zinc finger cores are predominantly screened by networks of backbone:core NH−S hydrogen bonds that electronically stabilize core thiolates and enhance backbone packing. Stabilizing protein:core interactions can be mapped to conserved residues, including [Arg,Lys]:core salt-bridges in some protein families. Labile zinc fingers are identified by poorly screened cores, possibly indicating redox or metallothionein (MT) regulated function. Consistent with experiment, the cores of the C-terminal finger of the human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein p7 (NCp7) and *Escherichia coli* Ada protein (Ada) "finger" are identified as reactive. The C-terminal zinc fingers of nuclear receptors are predicted to be the most labile in this study, particularly the human estrogen receptor (hER), which contains a triad of reactive thiolates. We propose that hER DNA binding is redox and MT regulated through the C-terminal finger and that *weak* electrophilic agents may inhibit hER-mediated transcription, implicated in breast cancer progression.

Introduction

Zinc fingers are small metal-binding peptide units that function in gene regulation by promoting specific protein−nucleic acid binding and recognition,[1,2] in some cases mediating protein−protein interactions as well.[3] All zinc fingers contain a core[4] of cysteine and histidine residues, tetrahedrally coordinated to Zn^{2+}, which anchors the peptide loops linked to these chelated residues. This design yields autonomous folding motifs, $X_i-\Omega-X_j-\Omega-X_k-\Omega-X_l-\Omega-X_m$, where Ω are Zn^{2+} ligands (Ω = Cys, His) and $\{X_i\}$ are peptide segments of length *i*. These protein domains are highly conserved, particularly the Zn^{2+}-coordination motif of the core, $Cys_\alpha His_\beta$,[4] and the core's residue sequence. The zinc finger cores are crucial to the stability and arrangement of local protein secondary structure, with numerous experiments demonstrating that Zn^{2+}-coordination is necessary for protein function. Conserved residues in loops $\{X_i\}$ also contribute to the stability of the zinc finger fold, as well as target-binding specificity.[2,5,6] Zinc fingers often occur in tandem as DNA-binding arrays, particularly in Cys_2His_2-containing transcription factors (TFs), where multiple fingers interact sequentially with the major[2,5,6] and minor[7] groves of DNA. The modular properties of zinc fingers, combined with the amino acid variability of $\{X_i\}$, are well suited for the combinatorial development of zinc finger designs that confer high DNA specificity.[2,5,6] These properties may also account for the abundance of encoded zinc fingers, estimated to be 1% of the human genome.[8] At least 10 functionally distinct classes of zinc fingers are categorized that utilize a variety of different cores and peptide loop topologies.[9]

In contrast to a variety of *enzymatic* zinc cores, where Zn^{2+} chemically functions as a Lewis acid in catalysis,[10−12] zinc finger cores are usually regarded as *structural* elements within proteins. However, zinc finger cores can be chemically active since their nucleophilic Cys thiolates are potentially vulnerable to oxidation, as observed for zinc fingers containing Cys_2His_2,[13−17] Cys_3-

* Corresponding authors: Andrew Maynard and David Covell, Screening Technologies Branch, Laboratory of Computational Technology, NCI-FCRDC, Bldg. 430, Room 215, Frederick, MD 21702. E-mail: covell@faxpdc.ncifcrf.gov.

† Present address: Astra Zeneca Inc., 1800 Concord Pike, NLW 3031, Wilmington, DE 19850. E-mail: Andrew.Maynard@astrazeneca.com.

(1) Berg, J. M.; Godwin, H. A. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 357−371.

(2) Choo, Y.; Klug, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 117−125.

(3) Mackay, J. P.; Crossley, M. *Trends Biochem. Sci.* **1998**, *23*, 1−4.

(4) In this paper, a zinc finger "core" is explicitly defined as Zn^{2+} and the coordinated side chains of Cys or His. The nomenclature adopted for the cores is based on the Zn^{2+}-coordination motif, $Cys_\alpha His_\beta$ ($\alpha + \beta = 4$, $2 < \alpha < 4$, $0 < \beta < 2$), which is independent of the residue sequence. Thus the core chemical composition can be expressed as $[ZnCys_\alpha His_\beta]^{(2-\alpha)}$, assuming a core charge of $(2 - \alpha)$.

(5) Berg, J. M.; Shi, Y. *Science* **1996**, *271*, 1081−1085.

(6) Wolfe, S.; Greisman, H.; Ramm, E.; Pabo C. *J. Mol. Biol.* **1999**, *285*, 1917−1934.

(7) Neely, L.; Trauger, J.; Baird, E.; Dervan, P.; Gottesfeld, J. *J. Mol. Biol.* **1997**, *274*, 439−445.

(8) Hoovers, J.; Mannens, M.; Bliek, J.; van Heyningen, V.; Porteous, D.; Leschot, N.; Westerveld, A.; Little, P. *Genomics* **1992**, *12*, 254−263.

(9) Schwabe, J.; Klug, A. *Nat. Struct. Biol.* **1994**, *1*, 345−349.

(10) Lipscomb, W.; Strater, N. *Chem. Rev.* **1996**, *96*, 2375−2433.

(11) Christianson, D. W. *Adv. Protein Chem.* **1991**, *42*, 281−355.

(12) Vallee, B. L.; Auld, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 220−224.

(13) Huang, M.; Maynard, A.; Turpin, J.; Graham, L.; Janini, G.; Covell, D.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 1371−1381.

10.1021/ja0011616 CCC: $20.00
Published on Web 01/20/2001

His,[13,16,18−23] and Cys_4 cores.[13,24−27] Increasing evidence suggests that some zinc fingers are predisposed to oxidation, possibly as a means of redox or metallothionein (MT) regulation of their function, while others are relatively inert, having fixed structural functions.[14,15,17,28,29] Moreover, the primary function of some zinc finger-like cores actually involves specific delivery of key thiolates for reaction with substrate. Examples include the *Escherichia coli* Ada protein, where repair of methylated DNA is accomplished by specific methyl transfer from phosphotriester to Cys^{69} of a Cys_4 core,[30] and more recently in *E. coli* methionine synthase, where homocysteine is part of a Cys_4 core that is catalytically methylated to methionine.[31] In these cases, the Cys_4 core is utilized as a reactive construct, representing a nucleophilic counterpart to enzymatic (electrophilic) zinc cores. It is interesting to speculate if these reaction mechanisms evolved from early zinc finger-like proteins containing labile cores.

The reactivity of zinc finger cores has been exploited in the pursuit of alternative retroviral therapies. A variety of electrophilic agents cause HIV-1 inhibition by chemical modification of thiolates in the NCp7 Cys_3His cores.[13,18,19,23] In the context of density-functional theory (DFT), the reactivity of these cores has been correlated with the "electrophilic power" of inhibitors[20,32] and indicates that Cys_3His cores are particularly vulnerable to soft electrophiles.[20] Although each of the two conserved NCp7 zinc fingers share the same retroviral zinc finger motif, the core of the C-terminal finger is substantially more reactive than the N-terminal finger,[21−23] with experiment[21−22] and theory[20] indicating that Cys^{49} is the most labile site of NCp7. Additional experiments demonstrate that some agents selectively react with the NCp7 zinc fingers, without affecting other cellular zinc finger proteins, containing Cys_2His_2, Cys_3His, and Cys_4 cores.[13] Although it remains unclear if HIV-1 NCp7 can be specifically targeted, these results illustrate that the reactivity of zinc fingers is varied, suprisingly, even for similar zinc finger motifs.

Both the chemical composition of a zinc finger core and its local protein environment are evident determinants of reactivity.[13,33] Cys_4 cores are expected to be the most nucleophilic, if one assumes a net anionic charge (−2), followed by Cys_3His (−1), and then neutral Cys_2His_2 (0). Kinetic studies of model zinc finger cores, composed of mixed complexes of phenylthiolate (PhS^-) and 1-metheylimidazole (MeIm), show that the rate of methylation does in fact follow the trend: $[Zn(PhS)_4]^{2-} > [Zn(PhS)_3(MeIm)]^- \gg [Zn(PhS)_2(MeIm)_2]$, and reactivity of the tetrathiolate center is nearly equal to free thiolate.[33] It is likely that the reactivity of zinc finger cores is systematically higher than these model cores, since alkylthiolates (Cys^-) are stronger nucleophiles. However, "bare" cores are not found in nature. Thus, the reactivity of a given core must be evaluated on an individual basis, in the context of its surrounding protein environment. Within proteins, amide backbone:core NH−S hydrogen bonds are often observed to electronically stabilize core thiolates, even in the case of neutral Cys_2His_2 cores.[34−36] This network of NH−S bonds is also important for determining the local protein fold, as well as its overall stability.[36,37] Positively charged Arg and Lys side chains can also form ionic pairs with anionic cores,[38] similar to salt-bridges. Notably, the Cys protonation state of zinc cores, under physiological conditions, remains unresolved. Although the above results, including high-resolution X-ray crystallographic[39] and EXAFS measurements[40,41] of Zn−S bond distances, which are very sensitive to the Cys protonation state,[42] are consistent with unprotonated Cys_4 and Cys_3His cores, recent mass spectrometric measurements indicate protonated Cys_4 and Cys_3His cores, such that these cores are charge-neutral.[43]

In this study the protein screening of zinc cores is utilized as a rationale for evaluating the chemical stability of zinc fingers and related zinc protein constructs. Given that cores with the same Zn^{2+}-coordination motif are chemically equivalent, their differential reactivities must be modulated by variation in the surrounding protein. We postulate that *structural* zinc finger motifs have evolved in a manner that has promoted steric and electronic shielding of their potentially reactive core thiolates, particularly in the case of anionic cores. Therefore, zinc fingers that utilize increasingly anionic (reactive) core motifs may exhibit a concomitant increase in the protein screening of their cores, which can be viewed as a consequence of the preservation of electroneutrality. In contrast, anionic cores lacking protein screening may signify zinc fingers particularly vulnerable to oxidation, possibly because their functionality is redox or MT

(14) Wu, X.; Bishopric, N.; Discher, D.; Murphy, B.; Webster, K. *Mol. Cell. Biol.* **1996**, *16*, 1035−1046.
(15) Xiaohua, C.; Chu, M.; Giedroc, D. *Biochemistry* **1999**, *38*, 12915−12925.
(16) Roehm, P. C.; Berg, J. M. *J. Am. Chem. Soc.* **1998**, *120*, 13083−13087.
(17) Zeng, J.; Vallee, B. L.; Kagi, J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9984−9988.
(18) Rice, W. G.; Schaeffer C.; Harten, B.; Villinger, F.; South, T.; Summers, M.; Henderson, L.; Bess, J.; Arthur, L.; McDougal, J.; Orloff, S.; Mendeleyev, J.; Kun, E. *Nature* **1993**, *361*, 473−475.
(19) Turpin, J. A.; Song, Y.; Inman, J.; Huang, M.; Wallqvist, A.; Maynard, A.; Covell, D.; Rice, W.; Appella, E. *J. Med. Chem.* **1999**, *42*, 67−86.
(20) Maynard, A. T.; Huang, M.; Rice, W. G.; Covell, D. G. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11578−11583.
(21) Chertova, E. N.; Kane, B.; McGrath, C.; Johnson, D.; Sowder, R.; Arthur, L.; Henderson, L. *Biochemistry* **1998**, *37*, 17890−17897.
(22) Hathout, Y.; Fabris, D.; Han, M.; Sowder, R.; Henderson, L.; Fenselau, C. *Drug Metab. Dispos.* **1996**, *24*, 1395−1400.
(23) Tummino, P. J.; Scholten, J.; Harvey, P.; Holler, T.; Maloney, L.; Gogliotti, R.; Domagala, J.; Hupe, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 969−973.
(24) Beerheide, W.; Bernard, H.; Tan, Y.; Ganesan, A.; Rice, W.; Ting, A. *J. Natl. Cancer Inst.* **1999**, *91*, 1211−1220.
(25) Casadevall, M.; Sarkar, B *J. Inorg. Biochem.* **1998**, *71*, 147−152.
(26) Cano-Gauci, D. F.; Sarkar, B. *FEBS Lett.* **1996**, *386*, 1−4.
(27) Hutchison, K.; Matic, G.; Meshinichi, S.; Bresnick, E.; Pratt, W. *J. Biol. Chem.* **1991**, *266*, 10505−10509.
(28) Maret, W.; Vallee, B. L. *Proc. Nat. Acad. Sci. U.S.A.* **1998**, *95*, 3478−3482.
(29) Andrews, G. K. *Biochem. Pharm.* **2000**, *59*, 95−104.
(30) Myers, L. C.; Wagner, G.; Verdine, G. L. *J. Am. Chem. Soc.* **1995**, *117*, 10749−10750.
(31) Peariso, K.; Goulding, C.; Huang, S.; Matthews, R.; Penner-Hahn, J. *J. Am. Chem. Soc.* **1998**, *120*, 8410−8416.
(32) Parr, R. G.; Szentpaly, L.; Liu, S. *J. Am. Chem. Soc.* **1999**, *121*, 1922−1924.
(33) Wilker, J. J.; Lippard, S. J. *Inorg. Chem.* **1997**, *36*, 969−978.
(34) Adman, E.; Watenpaugh, K.; Jensen, L. *Proc. Nat. Acad. Sci. U.S.A.* **1974**, *72*, 4854−4858.
(35) Carter, C., Jr.; Kraut, J.; Freer, S.; Alden, R. *J. Biol. Chem.* **1974**, *249*, 6339−6346.
(36) Blake, P. R.; Summers, M. F. *Advances in Inorganic Biochemistry*; Prentice Hall: NJ, 1994; pp 201−229.
(37) Berg, J. M. *Acc. Chem. Res.* **1995**, *28*, 14−19.
(38) Konrat, R.; Weiskirchen, R.; Bister, K.; Krautler, B. *J. Am. Chem. Soc.* **1998**, *120*, 7127−7128.
(39) Dauter, Z.; Wilson, K.; Sieker, L.; Moulis, J.; Meyer, J. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 8836−8840.
(40) Chance, M.; Sagi, I.; Wirt, M.; Frisbie, S.; Scheuring, E.; Chen, E.; Bess, J.; Henderson, L.; Arthur, L.; South, T.; Perez-Alvarado, G.; Summers, M *Proc. Nat. Acad. Sci. U.S.A.* **1992**, *89*, 10041−10045.
(41) Clark-Baldwin, K.; Tierney, D.; Govindaswamy, N.; Gruff, E.; Kim, C.; Berg, J.; Koch, S.; Penner-Hahn, J. *J. Am. Chem. Soc.* **1998**, *120*, 8401−8409.
(42) Ryde, U. *Euro. Biophys. J.* **1996**, *24*, 213−221.
(43) Fabris, D.; Hathout, Y.; Fenselau, C. *Inorg. Chem.* **1999**, *38*, 1322−1325.

Table 1. Database of Zinc Finger Protein Structures

core motif	protein family	PDB structure
Cys_2His_2	yeast Adr1 TF	1paa 2adr
	enhancer binding protein	3znf 5znf
	GAGA TF	1yui 1yuj
	HIV integrase	1aub 1wja 1wjc
	uridylyltransferase	1gup 1hxp
	TFIIIA	1tf3 1tf6
	Zif268	1a1f 1a1g 1a1h 1a1i 1a1j 1a1k 1a1l 1aay 1mey 1zaa
	miscellaneous	1ncs 1rmd 1sp2 1ubd 1zfd 1znf 1znm
Cys_3His	adenovirus DNA-binding protein	1adt 1adu 1adv 1anv
	endonuclease I−PpoI	1a73 1a74
	protein kinase C	1faq 1ptq 1tbn 1tbo
	HIV-1 nucleocapsid p7	1aaf 1bj6 1mfs
	Cys- and Gly- rich protein	1ctl 1cxx 1iml 1qli 1zfo
	RING-finger proteins	1bor 1chc
	tRNA-guanine transglycosylase	1pud 1wke 1wkf
	miscellaneous	1a7i 1a8h 1rmd 1ycs
Cys_4	alcohol dehydrogenase	1axg 1a71 1a72 1agn 1axe 1lde 1ldy 1teh 3bto
	adenylate kinase	1zin 1zio 1zip
	GATA TF	2gat 3gat 4gat 5gat 6gat 7gat
	rubredoxin	1irn 1zrp
	nuclear hormone receptor	1a6y 1gdc 1glu 1hcp 1hcq 1hra 1lat 1rxr 2nll
	Cys- and Gly- rich protein	1a7i 1qli 1ctl 1cxx 1iml
	RING-finger proteins	1bor 1chc
	polymerases	1a5t 1qyp
	adenovirus DNA-binding protein	1adt 1adu 1adv 1anv
	miscellaneous	1adn 1mea 1nbe 1pft 1tfi

regulated. To investigate these suppositions, we have evaluated the protein packing and electrostatic screening of 207 zinc finger cores (78 Cys_2His_2, 42 Cys_3His, 87 Cys_4), derived from 92 zinc finger protein structures in the Protein Data Bank.[44] These properties serve as steric and electronic measures of the protein protection (stabilization) of zinc cores and are used to predictively rank the relative stability of zinc fingers. We emphasize that our analysis is independent of a priori assumptions of the core protonation state(s), since only the protein environment of zinc cores is measured. We find the average protein screening of zinc finger cores is strongly differentiated by $Cys_\alpha His_\beta$, observing a concomitant increase in the screening of anionic motifs: $Cys_4 \gg Cys_3His > Cys_2His_2$, consistent with $[ZnCys_3His]^{(-1)}$ and $[ZnCys_4]^{(-2)}$ cores, based on the preservation of electroneutrality. Interestingly, a wide diversity of core environments is also observed. Notably, the C-terminal finger of nuclear hormone receptor (NR) DNA-binding domains (DBDs) is predicted to be the most labile zinc finger motif in this study, and within the NR family, the human estrogen receptor (hER) is calculated to have the most labile core. We propose the C-terminal hER zinc finger functions as a redox or MT regulated "switch" that controls hER DNA binding and that *weak* electrophiles may disrupt this finger, thus inhibiting hER-mediated transcription.

Methods

(a) Construction of the Zinc Finger Database. The database of zinc finger proteins was constructed from experimental X-ray crystallographic and NMR structures of the Protein Data Bank (PDB).[44] The criteria for selection was based on proteins containing at least one 4-coordinate Zn site, consisting only of Cys or His, having Zn−S and Zn−N distances within 2.6 Å and 2.3 Å, respectively. To uniformly treat PDB structures that did or did not include hydrogen atoms, all hydrogen positions were assigned with the *protonate* utility of AMBER5.0.1.[45] Zn-bound Cys was treated as Cys^-, but this did not effect the protein screening analysis (see below). Water, nucleic acid, or complexed ligands were also excluded from analysis, since the protein environment of the cores was the fundamental property of interest. To eliminate statistical bias in the dataset due to multiple PDB entries of a given protein, averages were computed over protein families, with equal weight assigned to each family. Thus, the average property (P) of a given zinc finger core motif was

$$P_{\text{motif}} = \sum_i^{N_f} \bar{P}_i / N_f \tag{1}$$

where $\bar{P}_i$ is the average over all structures within family i and N_f is the number of families containing a given $Cys_\alpha His_\beta$ core. Calculation of $\bar{P}_i$ also provided a useful internal measure of the sensitivity of P to structural variation in the PDB data. Proteins having only a single structure were grouped as a "miscellaneous" family. The database of structures and protein families are listed in Table 1. Zinc clusters or binuclear zinc fingers, as found in MT and the GAL4 family, were not included in this study.

(b) Evaluation of Protein Packing about Zinc Finger Cores. Calculation of the protein density was based on the approach of Grant et al.,[46] where the molecular shape-density of a protein (ρ^g) is defined as a sum of atom-centered, spherical Gaussians,

$$\rho^g(\mathbf{r}) = \sum_i \rho_i - \sum_{i<j} \rho_i \rho_j + \sum_{i<j<k} \rho_i \rho_j \rho_k - \cdots \tag{2}$$

where,

$$\rho_i(r_i) = w_i \, exp(-\alpha_i^2 r_i^2) \tag{3}$$

In eq 3, r_i is a local coordinate, $r_i = |\mathbf{r} - \mathbf{r}_i|$, the Gaussian width ($1/\alpha_i$) is proportional to the atomic van der Waals radius, and the weight (w_i)

(44) Bernstein, F.; Koetzle, T.; Williams, G.; Meyer, E.; Brice, M.; Rodgers, J.; Kennard, O.; Shimanuchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535−542.

(45) Pearlman, D.; Case D.; Caldwell, J.; Cheatham T.; Ross, W.; Simmerling, C.; Darden, T.; Merz, K.; Stanton, R,; Cheng, A.; Vincent, J.; Crowley, M.; Radmer, R.; Ferguson, G.; Seibel, G.; Singh, U.; Weiner, P.; Kollman, P. *AMBER 5.0.1*; University of California: San Francisco, 1998.

(46) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. *J. Comput. Chem.* **1996**, *17*, 1653−1666.

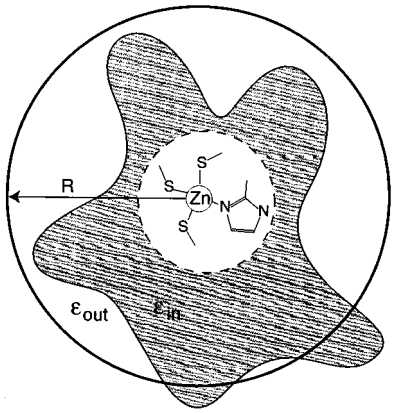

Figure 1. Schematic protein environment of a zinc finger core, defined by Zn^{2+} and the coordinated residue side chains, Cys_3His in this example. The core is embedded in a protein (dielectric constant = ϵ_{in}), surrounded by a solvent continuum (dielectric constant = ϵ_{out}). Protein screening of the core is evaluated within successive radii R.

is chosen so that the van der Waals volume is conserved, $v_i = \int d\mathbf{r}_i \rho_i$. The protein volume, $V^g = \int d\mathbf{r} \rho^g$, can be efficiently computed by virtue of analytical Gaussian integrals. As previously demonstrated,[46] truncation of eq 2 to six-body terms (overlaps) yielded protein volumes in agreement with hard-sphere calculations to within 0.3%. The van der Waals radii (Å) for H, C, N, O, S, and Zn were 1.1, 1.7, 1.65, 1.6, 1.9, and 1.4, respectively.

The packing fraction or density of the protein, surrounding a zinc finger core, was defined as the proportion of molecular volume within a sphere of radius R, $\rho(R) = 3V^g(R)/4\pi R^3$, centered at the core, as illustrated in Figure 1. ρ was evaluated with the origin located at the $Zn(\rho^{Zn})$ and $S(\rho^S)$ positions. Since the steric profile of cores with the same Zn^{2+}-coordination motif is radially uniform, all core atoms (Zn, Cys, and His side chains) were excluded from the evaluation of ρ. Thus, evaluation of ρ was independent of core motif and measured only the radial distribution of the surrounding protein density. ρ was also decomposed into contributions from the peptide backbone (ρ_{BB}), backbone NH groups (ρ_{NH}) and hydrophobic groups (ρ_{phobe}: side chains of Gly, Ala, Val, Leu, Ile, Pro, Trp, Tyr, and Phe), by *only* evaluating the density of these respective atom sets. As a benchmark, the $\rho(R)$ profile of the GATA-1 core (PDB: 2gat), computed to 16 Å in 0.5 Å steps, required 13s of CPU time on a SGI R12000 processor.

(c) Electrostatic Screening Analysis of Zinc Finger Cores. Electrostatic screening of zinc finger cores was evaluated with the program DelPhi (version 2.50, InsightII 98.0, MSI Inc., San Diego, CA), developed by Honig and co-workers.[47] All atoms within 30 Å of the Zn position defined the dielectric boundary between protein and solvent. This encompassed the entire protein volume for most proteins studied and our results were insensitive to this cutoff. As illustrated in Figure 1, the radial dependence of the electrostatic potential applied at a zinc finger core, due to the surrounding protein, $V_s(R)$, was determined by charging successively larger spheres of protein atoms and solving Poisson's equation. In this treatment, only atoms within a radius R of the Zn position carried their full charge, while charges of atoms outside this range were "switched" off. For a given R, the total protein electrostatic potential, $V_{protein}$, was calculated by solving Poisson's equation on a large, coarse grid (grid spacing = 0.9 Å, protein dimension <70% of grid dimension, full Coulombic boundary conditions). This grid provided boundary and initial conditions for solving Poisson's equation on a focused grid (grid spacing = 0.6 Å, protein dimension <95% of grid dimension). The identical procedure was used to calculate the core electrostatic potential, V_{core}, with *only* the core atoms charged. The protein screening potential of the core was thus, $V_s = V_{protein} - V_{core}$, with V_s evaluated at Zn and S positions. Refinement of the grid density to 0.4 Å did not significantly change (±1%) V_s in test cases. V_s was decomposed into protein contributions from the total backbone, V_s^{BB}, side chain, V_s^{SC}, and side chain subgroups, by charging only these subsets of atoms. As a benchmark, the $V_s(R)$ profile of the GATA-1 core (PDB: 2gat), computed to 25 Å in 1 Å steps, required 4 min of CPU time on a SGI R12000 processor.

Dielectric constants of the protein and solvent were $\epsilon_{in} = 2$ and $\epsilon_{out} = 80$, respectively. The van der Waals radii were identical to the packing analysis described above and the solvent probe radius was 1.4 Å. Both AMBER[45] and CHARMM[48] charge sets were used. Zn-bound cysteines, treated as Cys^-, required additional parametrization, since Zn−S bonds have covalent character. Zn and Cys charges were based on DFT calculations,[20,49] using Dmol (version 960 MSI Inc., San Diego, CA) to compute optimized core geometries and Hirshfeld and electrostatic potential fitted charges. Zn was assigned a partial charge of $q_{Zn} = +1$, while charges for S and C^β of Cys^- side chains, (q_S,$q_{C\beta}$), were (−0.34,−0.3), (−0.473,−0.333), and (−0.54,−0.35) for the Cys_2His_2, Cys_3His, and Cys_4 motifs, respectively. These charges obey empirical formulas, $q_S = q_S^0 + \delta_S$ and $q_{C\beta} = q_{C\beta}^0 + \delta_{C\beta}$, where $q_S^0 = -0.74$ and $q_{C\beta}^0 = -0.4$, correspond to charges of unbound Cys^-. Charge transfer to Zn was modeled as $\delta_S + \delta_{C\beta} = (2 - q_{Zn})/N_{cys}$, such that $\delta_S/\delta_{C\beta} = 4$, where N_{cys} is the number of Zn-bound cysteines; the charge-transfer per Cys^- is inversely proportional to N_{cys} and the majority of the charge originates from S. The $H^{C\beta}$ charge was 0.09, and C^α charges were −0.002 and 0.06 for the AMBER and CHARMM sets, respectively. All other Cys charges were fixed to their respective force fields, yielding net core charges of 0, −1, and −2 for Cys_2His_2, Cys_3His, and Cys_4, respectively. It is important to note that V_s was insensitive to the modeling of core partial charges, due to cancellation of V_{core}, $V_s = V_{protein} - V_{core}$. A random ±25% variation in the core partial charges resulted in less than 0.01% variation in $V_s(R)$ and treating Cys_3His and Cys_4 cores as charge-neutral ($q_i^{core} \equiv 0$) resulted in only a 0.4% variation in $V_s(R)$.

(d) Analysis of Variance in Packing and Screening. Statistical differences between packing and screening profiles, as a function of $Cys_\alpha His_\beta$, were evaluated by balanced single-factor analysis of variance (ANOVA),[50] using MATLAB (version 5.3.1, The MathWorks, Inc., Natick, MA). At a given R, a population of PDB observations exists for each independent variable (core motif). ANOVA provided a statistical measure of the similarity in these population distributions, based on the likelihood (p-value, p) that different populations share a common mean (null hypothesis). Low p, typically $p \leq 0.05$, corresponds to significantly different $Cys_\alpha His_\beta$ populations. Over a range R, the median ANOVA p-value ($p_{1/2}$) of two profiles was a robust statistic of profile similarity. Since steric and electrostatic screening of zinc cores are local and long-range properties, respectively, an 8 Å cutoff was used for packing profile ANOVA, while full profiles were used for electrostactic screening ANOVA.

Results and Discussion

(a) Trends in the Protein Packing of Zinc Finger Cores. The steric screening of zinc finger cores is measured by the protein packing density, ρ. Radial profiles of ρ are shown in Figure 2, with each profile corresponding to the average trend in ρ^{Zn} for a given core motif. Similar results are obtained for average densities centered at the core thiolate positions, ρ^S. It is important to note the different trends observed in short-range structuring of ρ as a function of $Cys_\alpha His_\beta$, namely the position of the shoulder, as well as the magnitude and sharpness of the maximum density, ρ_{max}. Figure 2a illustrates that Cys_4 cores are on average the most tightly packed, followed by Cys_3His cores, and finally Cys_2His_2 cores: $\rho^{C_4} > \rho^{C_3H} > \rho^{C_2H_2}$. The values of ρ_{max} for Cys_4 and Cys_3His cores are 19% greater than that of Cys_2His_2. Statistically, the ANOVA similarity of $\rho^{C_4}(R)$, compared to $\rho^{C_3H}(R)$ ($p_{1/2} = 0.02$) and $\rho^{C_2H_2}(R)$ ($p_{1/2} = 4 \times$

(47) Honig, B.; Sharp, K.; Yang, A. *J. Phys. Chem.* **1993**, *97*, 1101−1109.

(48) MacKerell, A.; Bashford, D.; Dunbrack, R.; Evanseck, J.; Field, M.; Fisher, S.; Gao, J.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, I.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102* 3586−3616.

(49) Maynard, A. T.; Covell, D. G. Manuscript in preparation.

(50) Neter, J.; Wasserman, W.; Kunter, M. *Applied Linear Statistical Models*; Homewood: IL, 1985; p 517.

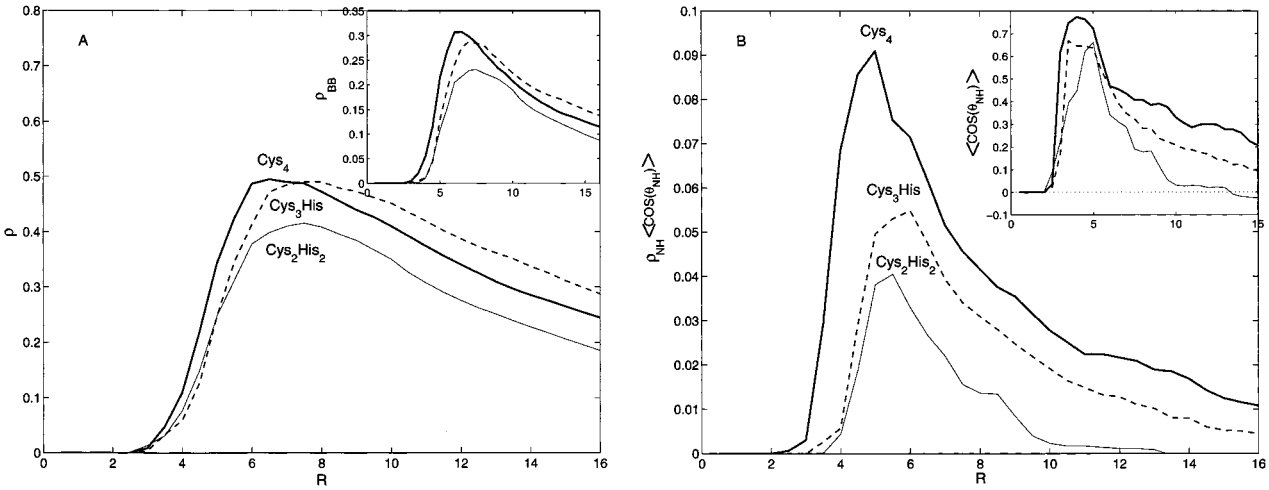

Figure 2. Average radial (Å) protein packing profiles of zinc finger cores: Cys_4 (bold), Cys_3His (dashed), Cys_2His_2 (light solid). (**A**) Total protein density, ρ, and total backbone density, ρ_{BB} (inset). Standard deviation in ρ is 0.03, 0.06, and 0.03 for Cys_4, Cys_3His, and Cys_2His_2, respectively. Standard deviation in ρ_{BB} is 0.03, 0.04, and 0.02 for Cys_4, Cys_3His, and Cys_2His_2, respectively. (**B**) Total backbone NH density, ρ_{NH}, weighted by the $\overrightarrow{NH}$ orientation toward the core Zn position (inset), $\langle\cos(\theta_{NH})\rangle$, $\theta_{NH} = \angle H{-}N{-}Zn$. Standard deviation in $\rho_{NH}\langle\cos(\theta_{NH})\rangle$ is 0.004, 0.006, and 0.003 for Cys_4, Cys_3His, and Cys_2His_2, respectively.

10^{-5}), is well separated. Although there is significant similarity between ρ^{C_3H} and $\rho^{C_2H_2}$ ($p_{1/2} = 0.1$), at ρ_{max} packing is strongly differentiated ($p = 0.004$). Decomposition of ρ reveals that the trend in total packing density is correlated with the backbone packing component of ρ, ρ_{BB}, shown in the inset of Figure 2a. In addition, the contrast between ρ_{BB} profiles of the different core motifs is more pronounced than ρ, and statistically more distinct. The backbone is much more tightly structured around Cys_4 cores ($\rho_{BB}[R_{max} = 6.5\ Å] = 0.31$) than the other motifs, becoming less so for Cys_2His_2 cores ($\rho_{BB}[R_{max} = 7.5\ Å] = 0.23$): $\rho_{BB}^{C_4} > \rho_{BB}^{C_3H} > \rho_{BB}^{C_2H_2}$. Contraction of the backbone around Cys-rich cores may partly be explained by the replacement of the more bulky His side chain (76 $Å^3$) with the smaller Cys side chain (44 $Å^3$). However, anionic core motifs have an enhanced network of backbone:core NH−S hydrogen bonds that likely stabilize tighter backbone folding around these cores. The average number of backbone NH−S hydrogen bonds per thiolate ($\#_{NH-S}$) increases as 1.26 (Cys_2His_2), 1.34 (Cys_3His), and 1.60 (Cys_4), with an overall NH−S hydrogen bond distance of 2.57 Å, based on a 2.8 Å NH−S cutoff. Thus, on average the number of backbone:core NH−S bonds per core increases as 2.5 (Cys_2His_2), 4.0 (Cys_3His), and 6.4 (Cys_4), with each Cys contributing a satellite of NH−S interactions. Together these results are consistent with complementary protein:core interactions that preserve electroneutrality, supporting an anionic trend in effective core charges: Cys_2His_2 (0), Cys_3His (−1), Cys_4 (−2).

The average trend in backbone NH packing density and orientation of the different core motifs is illustrated in Figure 2b. Here, the density of each NH unit is weighted by its $\overrightarrow{NH}$ direction cosine, relative to the center of the core, defined as the Zn position ($\theta_{NH} = \angle$ H−N−Zn). The maxima of $\rho_{NH}\langle\cos(\theta_{NH})\rangle$ for Cys_4 and Cys_3His cores is on average 124 and 36% higher than Cys_2His_2 cores, respectively. Statistically, the dissimilarity of $\rho_{NH}\langle\cos(\theta_{NH})\rangle$ profiles of Cys_4, compared to Cys_3His ($p_{1/2} = 2 \times 10^{-6}$) and Cys_2His_2 ($p_{1/2} = 1 \times 10^{-10}$), is extremely pronounced, as is Cys_3His compared to Cys_2His_2 ($p_{1/2} = 3 \times 10^{-4}$). The inset of Figure 2b shows the local orientation of NH groups tends to be more ordered around increasingly anionic cores. Moreover, the trend in $\langle\cos(\theta_{NH})\rangle$ indicates enhanced long-range backbone ordering for negative cores, particularly Cys_4 cores, extending beyond the range of NH−S interactions. This structuring may be due to increased backbone:backbone NH−O interactions in the vicinity of anionic cores. We postulate that the more extensive network of NH−S backbone:core interactions, found for anionic cores, may induce outlying NH−O hydrogen bonding, given the local bias in the amide dipole orientation around these cores. This is supported by the computed distribution of NH−O bonds emanating from the vicinity of zinc finger cores, which is broadly peaked between 8 and 13 Å (figure not shown). On the basis of the maxima of this distribution, the number of NH−O bonds surrounding Cys_4 and Cys_3His cores is on average 40 and 14% higher than Cys_2His_2 cores, respectively. Consequently, these results imply that mutation or chemical modification of core cysteines would drastically degrade the native protein structure by disrupting the interior scaffold of NH−S bonds and coupled NH−O interactions.

The average hydrophobic packing, ρ_{phobe}, of neutral Cys_2His_2 cores is substantially tighter and more pronounced than Cys_3His or Cys_4 motifs ($\rho_{phobe}^{C_2H_2} > \rho_{phobe}^{C_3H} > \rho_{phobe}^{C_4}$), *opposite* the trends observed for ρ_{BB} and ρ. This suggests that Cys_2His_2 cores are more conducive to formation of zinc finger "hydrophobic cores". Conserved aromatic residues, located in the vicinity of Cys_2His_2 cores, are known to confer zinc finger stability, as well as enhanced metal ion binding affinity.[1] The Cys_4 profile is very slightly more hydrophobic than Cys_3His, but this result is skewed by the extraordinarily high ρ_{phobe} profile of the Cys_4 rubredoxin core (2 standard deviations above $\rho_{phobe}^{C_4}$). Although CH−S hydrogen bonds are substantially weaker than NH−S interactions, scalar coupling between aliphatic side chains and Fe Cys_4 cores has been proposed as a possible electron-transfer mechanism in rubredoxins.[36] If rubredoxin is excluded from the data, the statistical dissimilarity between ρ_{phobe} profiles of Cys_2His_2, compared to Cys_3His ($p_{1/2} = 0.04$) and Cys_4 ($p_{1/2} = 0.003$), is well separated. In contrast, the ρ_{phobe} profiles of Cys_4 and Cys_3His are quite similar ($p_{1/2} = 0.4$).

(b) Electrostatic Screening Profiles of Zinc Finger Cores. Decomposition of ρ indirectly reveals that the electronic environment of zinc finger cores depends strongly on the Zn^{2+}-coordination motif. Continuum electrostatic calculations are used to explicitly probe the electrostatic screening of zinc finger cores. Figure 3 shows the average radial dependence of the electrostatic potential exerted by the surrounding protein on a core, $V_s(R)$,

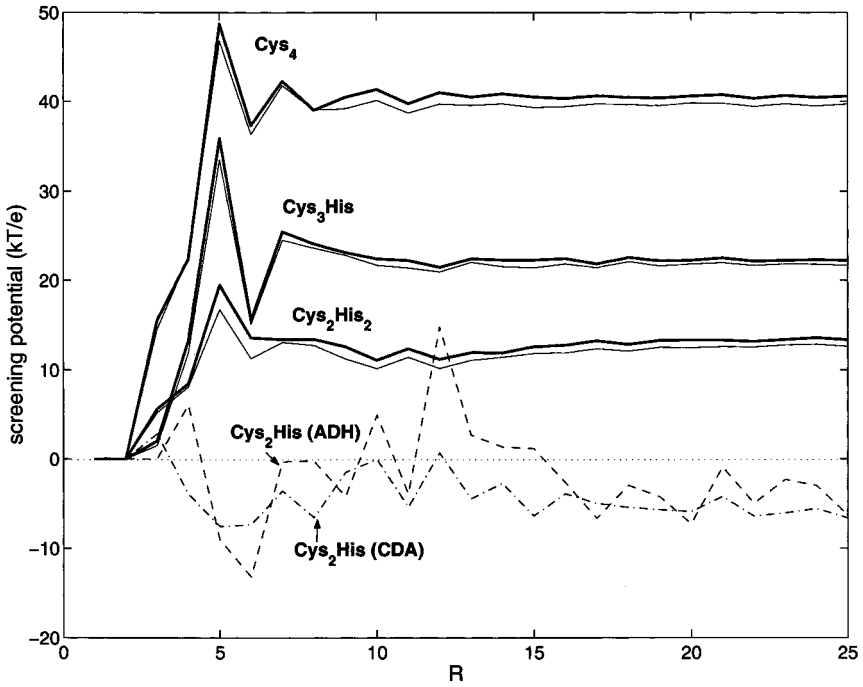

Figure 3. Average radial (Å) profiles of the protein electrostatic screening potential (kT/e) of zinc finger cores, $V_s(R)$: AMBER charge set (bold), CHARMM charge set (light solid). V_s profiles of the enzymatic Cys_2His cores of alcohol dehydrogenase, ADH (dash dot), and cytidine deaminase, CDA (dash), are shown for comparison. Standard deviation in V_s is 6.5, 4.1, and 10.2 kT/e for Cys_4, Cys_3His, and Cys_2His_2, respectively.

as a function of $Cys_\alpha His_\beta$, with V_s centered at the Zn position. Similar profiles are also obtained by averaging V_s over the core S positions. The magnitude of V_s is inversely proportional to the protein dielectric constant, however the relative separation of V_s, as a function of $Cys_\alpha His_\beta$, is invariant to ϵ in the range $1 < \epsilon < 4$. Consistent with the packing analysis above, the trend in V_s follows a complementary protein:core charge−charge interaction: $V_s^{C_4} \gg V_s^{C_3H} > V_s^{C_2H_2}$; anionic core motifs tend to be much more positively screened by protein. This supports the likelihood that Cys_3His and Cys_4 cores are anionic in the protein native state (i.e., Cys is deprotonated), based on the preservation of electroneutrality. There is markedly stronger short-range electrostatic screening of Cys_3His and particularly Cys_4 motifs, over Cys_2His_2 motifs. On the basis of the maxima of V_s at 5 Å (Figure 3), the short-range electrostatic screening of Cys_4 and Cys_3His cores is on average 2.5 and 1.8 times greater than Cys_2His_2 cores, respectively. The maxima at 5 Å correspond to the satellite of backbone:core NH−S interactions. Asymptotically, the protein screening of the Cys_4 and Cys_3His cores is 3.1 and 1.7 times greater than Cys_2His_2 cores, respectively. Essentially the same results are obtained with either AMBER or CHARMM charge sets (Figure 3). Statistically, the dissimilarity of $V_s^{C_4}$, compared to $V_s^{C_3H}$ ($p_{1/2} = 1 \times 10^{-4}$) and $V_s^{C_2H_2}$ ($p_{1/2} = 1 \times 10^{-5}$), is very pronounced. There is less contrast between $V_s^{C_3H}$ and $V_s^{C_2H_2}$ ($p_{1/2} = 0.05$), due to a large variance in $V_s^{C_2H_2}$ (10 kT/e) in closer proximity to $V_s^{C_3H}$. The large variance in $V_s^{C_2H_2}$ is attributed to a lack of correlated charge screening for these charge-neutral cores. Although the *net* charge of Cys_2His_2 is zero, the thiolates partial negative charge is stabilized by the presence of NH−S hydrogen bonds ($\#_{NH-S} = 1.26$), which contribute to the observed positive screening of Cys_2His_2 cores (locally, $V_s^{C_2H_2} = 19$ kT/e). The protein:core electrostatic interaction energy, $\Sigma q_i^{core} V_s(\mathbf{r}_i)$, yields a larger differentiation of core motifs, assuming anionic Cys_3His and Cys_4 cores. However, this measure is sensitive to the partial charge parameters of the core, as well as assumptions about the Cys protonation state of anionic cores, and therefore was excluded. Our analysis circumvents these problems, since by definition V_s is insensitive to modeling of the cores.

To compare zinc finger and enzymatic cores, V_s was calculated for the enzymatic cores of cytidine deaminase (CDA, PDB: 1aln, 1ctt), and liver alcohol dehydrogenase (ADH, PDB: 3bto). Like zinc fingers, these zinc enzymatic cores are composed only of Cys and His residues, in a Cys_2His motif, the fourth coordination site occupied by H_2O or substrate. The average V_s profiles of the CDA and ADH cores, shown in Figure 3, have a completely different signature in comparison to zinc finger cores, indicating different functionality. In contrast to zinc finger cores, the CDA and ADH cores are *negatively* screened, due to the presence of nearby Asp and Glu residues which can polarize the substrate (ADH) or directly activate H_2O (CDA). Furthermore, these enzymatic cores have fewer NH−S bonds ($\#_{NH-S} = 1.0$ CDA, 0.5 ADH), compared to zinc fingers. This latter feature has been postulated to enhance the covalent component of the Zn−S bond, thus providing a "valence buffer" of the catalytic site.[51] Recently, a novel Cys_3 cytidine deaminase has been found to have similar catalytic activity to its Cys_2His counterpart,[52] although Cys_3-coordination is more associated with zinc finger cores.

Decomposition of V_s into the total backbone, V_s^{BB}, and side chain, V_s^{SC}, contributions, is shown in Figure 4. The local profile of V_s is dominated by the backbone interaction with the core, particularly in the case of Cys_4 cores. The trend in V_s^{BB} varies as: $Cys_4 \gg Cys_3His > Cys_2His_2$, and is correlated with the number of backbone:core NH−S interactions. Based on the maxima of V_s^{BB} ($R = 5$ Å), the ratios of V_s^{BB} for the different motifs (Cys_4 to Cys_2His_2: 3.1, Cys_3His to Cys_2His_2: 1.9) are similar to ratios of the average number of NH−S bonds per core (Cys_4 to Cys_2His_2: 2.6, Cys_3His to Cys_2His_2: 1.6). At longer distances, the backbone becomes screened from the core and

(51) Xiang, S.; Short, S.; Wolfenden, R.; Carter, C., Jr. *Biochemistry* **1996**, *35*, 1335−1341.
(52) Carlow, D. C.; Carter, C., Jr.; Mejlhede, N.; Neuhard, J.; Wolfenden, R. *Biochemistry* **1999**, *38*, 12258−12265.

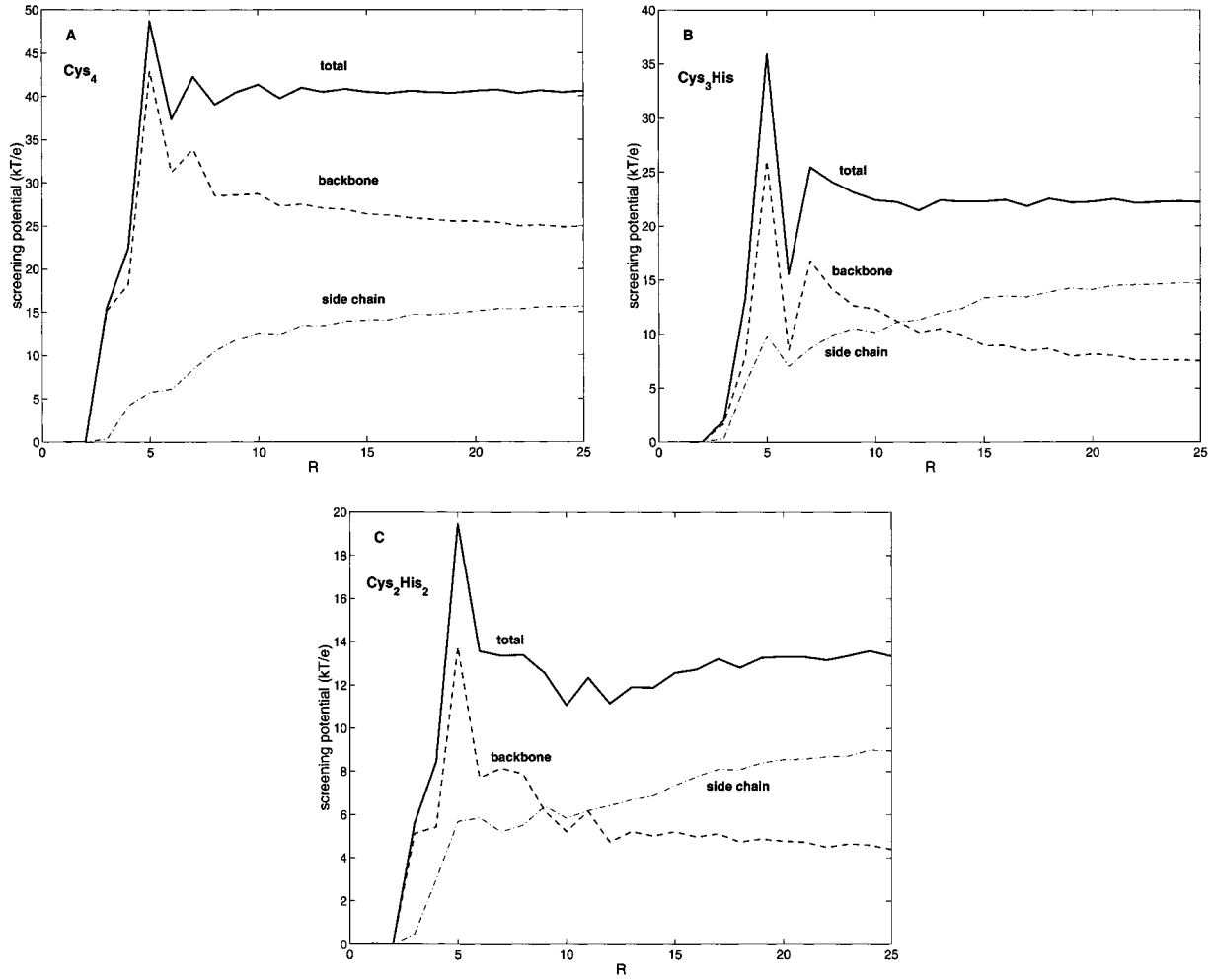

Figure 4. Decompostition of electrostatic screening profiles into total backbone, V_s^{BB}, and total side chain, V_s^{SC}, contributions: V_s (bold), V_s^{BB} (dash), V_s^{SC} (dash dot). (**A**) Decomposition of Cys_4 electrostatic screening. (**B**) Decomposition of Cys_3His electrostatic screening. (**C**) Decomposition of Cys_2His_2 electrostatic screening.

its structure less correlated with the proximity of the core, hence the decline in V_s^{BB}. However, long-range backbone screening of Cys_4 cores is more pronounced than the other motifs. As discussed above, this is attributed to induced backbone:backbone NH−O interactions in the vicinity of Cys_4 cores. Asymptotically, the ratio of V_s^{BB} for the different motifs varies as Cys_4 to Cys_2His_2: 5.7, Cys_3His to Cys_2His_2: 1.7. On average, the net side chain electrostatic screening of zinc finger cores is also positive. However, the profiles of V_s^{SC} are much less structured than V_s^{BB}. Asymptotically, V_s^{SC} becomes the dominant component of V_s, except for Cys_4 cores, its proportion of V_s increasing as 39, 66, and 68%, for Cys_4, Cys_3His, and Cys_2His_2 cores, respectively. Further decomposition of V_s^{SC} into component side chain contributions V_s^{pos} (Arg, Lys), V_s^{neg} (Asp, Glu), V_s^{pol} (Asn, Cys, Gln, His, Met, Ser, Thr), V_s^{ali} (Ala, Gly, Ile, Leu, Pro, Val), and V_s^{aro} (Phe, Trp, Tyr), reveals that the asymptotic trend in V_s^{SC} is most correlated with V_s^{pos}. This result is consistent with the abundance of Lys and Arg residues found in the zinc finger domains of TFs, where these residues often function in specific nucleic acid interactions. Notably, the combined influence of aliphatic, polar, and aromatic side chains (Cys_2His_2), generally yields a positive screening potential that cancels the negative contributions of Asp and Glu (V_s^{neg}).

In some zinc finger families, V_s^{pos} *locally* screens Cys_4 and Cys_3His cores, as shown in Figure 5. In these cases, individual Arg and Lys side chains directly interact with the thiolates of anionic cores, in a manner reminiscent of protein salt-bridges, contributing a large positive screening potential to the core. As examples, buried Arg interacts with the Cys_4 cores of NRDBDs in RevErb orphan NR ($Arg^{150,157}$), glucocorticoid NR ($Arg^{489,496}$), estrogen NR (Arg^{63}), and retinoid NR (Arg^{191}). Strong Arg:core interactions are also found for the Cys_4 core of *E. coli* DNA polymerase III (Arg^{46}), and the Cys_3His core of endonuclease I-PpoI (Arg^{122}). In the RING finger domain of the promyelocytic leukemia TF, there is a strong Lys^{20}:Cys_3His interaction. Arg and Lys side chains (Arg^{122}, $Lys^{152,174}$) are also closely distributed around the Cys_3His and Cys_4 cores of the cysteine- and glycine-rich protein family (CRP). However, in the case of CRP fingers these residues tend to be solvent-exposed, and their interaction with the cores is substantially weaker. Similarly, there are significant $Arg^{42,47}$:Cys_4 interactions observed for GATA-1, but these interactions are sensitive to structural variation within the NMR data. A recent study suggested that $Arg^{42,47}$ may impede the docking of electrophilic compounds into reactive proximity of the GATA-1 core.[13]

(c) Biochemical Diversity of Zinc Finger Core Environments: Assessing Reactivity/Stability of Individual Cores. Thus far we have compared average trends in the protein screening of zinc cores. Here the protein screening of cores is analyzed on an individual basis, revealing a wide diversity of core environments. This diversity may reflect differences in zinc finger functionality, where unprotected cores may signify redox

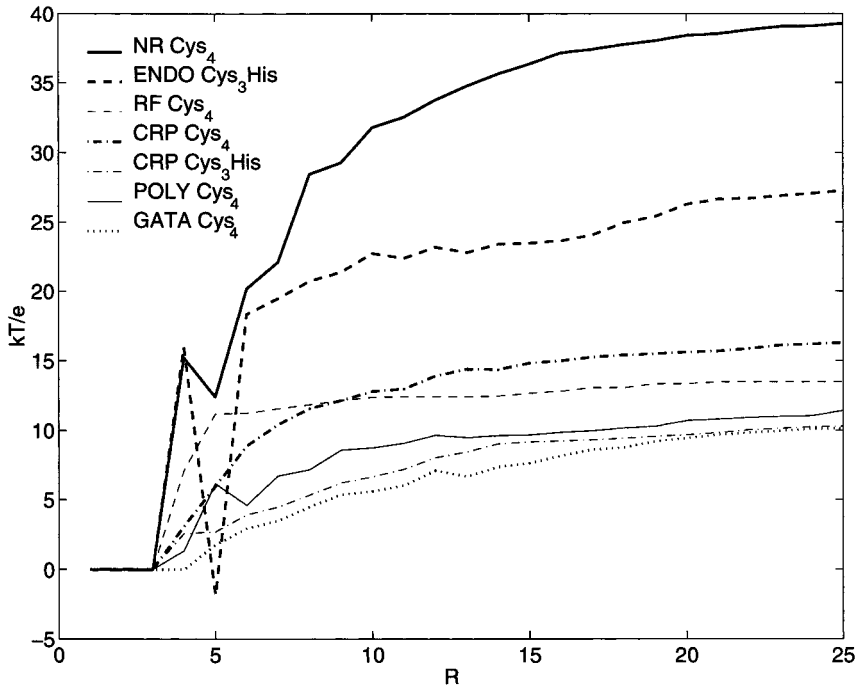

Figure 5. Electrostatic screening profiles of [Arg,Lys]:core interactions, $V_s^{pos}(R)$: nuclear hormone receptors (NR), endonuclease I-ppoI (ENDO), Cys-Gly rich proteins (CRP), RING-finger proteins (RING), RNA polymerase II (POLY), GATA TF family (GATA).

regulated fingers, while well-screened cores denote stable structural elements. Individual cores are ranked by steric (local) and electrostatic measures of their protein screening. The average packing density of a core, $\rho_s = \Sigma_i^{R_{max}} \rho(R_i)\Delta R/R_{max}$, within a radius R_{max}, is used as a steric screening parameter. R_{max} is based on the maxima of the average packing profiles (Figure 2A), $R_{max} \equiv 7$ Å, and ρ_s is proportional to the local integral of $\rho(R)$. Since the protein electrostatic potential can be long-range, the asymptotic value of V_s, $V_s^\infty \equiv V_s[R = 25$ Å$]$, is used as a measure of electrostatic screening. The sensitivity of ρ_s and V_s^∞ to structural variation within the PDB data was 4 and 15%, respectively, based on their standard deviations for multiple PDB entries of a given protein. Figure 6 is a composite representation of the steric and electrostatic environments of all zinc finger cores in this study. There is a tendency for ρ_s and V_s^∞ to be correlated, consistent with the previously discussed correlation between the average trends in ρ and V_s, where increased backbone packing of cores is associated with enhanced NH−S interactions. The overall shifts in the ρ_s and V_s^∞ distributions of Figure 6, $\{V_s^\infty,\rho_s\}^{motif}$, as a function of $Cys_\alpha His_\beta$, correspond to relative differences in the average trends of ρ and V_s. Figure 6 shows considerable overlap between $\{V_s^\infty,\rho_s\}^{motif}$ distributions. Thus, in some cases a Cys_4 core may be less screened than a Cys_2His_2 core, and is consequently predicted to be much more labile. Conversely, well separated regions of $\{V_s^\infty,\rho_s\}$ may indicate a protected Cys_4 core that is less reactive than a given Cys_2His_2 core.

Based on the experiments of Wilker et al.,[33] the reactivities of "bare" zinc finger cores are estimated to vary as 1000 (Cys_4):100 (Cys_3His):1 (Cys_2His_2). Recently, Huang et al.[13] also demonstrated that some electrophilic agents selectively modified NCp7 Cys_3His cores, thereby disrupting HIV replication, without inhibiting cellular zinc finger proteins containing Cys_4 (GATA-1) and Cys_2His_2 (Sp1) cores. Collectively, these results suggest that a core's protein environment is a key determinant of reactivity, as important as a core's intrinsic nucleophilic strength. Notably, the core of the N-terminal NCp7 finger is ranked near the average of $\{V_s^\infty,\rho_s\}^{Cys_3His}$ (Figure 6: $1aaf_1$), while the C-terminal NCp7 core is ranked as one of the least screened cores in $\{V_s^\infty,\rho_s\}^{Cys_3His}$ (Figure 6: $1aaf_2$). The V_s^∞ and ρ_s values of the N-terminal NCp7 core are 1.5 times greater than the C-terminal core, consistent with experiments demonstrating the N-terminal finger is substantially less reactive than the C-terminal.[21−23] The V_s^∞ and ρ_s values of the GATA-1 and Sp1 cores also fall in the middle of their respective distributions. The Cys_4 core of GATA-1 (Figure 6: 2gat) is 1.5 and 2.8 times more sterically and electrostatically screened, respectively, than the C-terminal NCp7 core. Experimentally, GATA-1 is much less vulnerable to electrophiles than NCp7,[13] although a Cys_4 core is expected to be an order of magnitude more reactive than a Cys_3His core.[13] In the case of Sp1 (Figure 6: 1sp2), its neutral Cys_2His_2 core is electrostatically *less* screened than the C-terminal core of NCp7. Thus, the inertness of Sp1[13] is likely an example of its intrinsically more stable Cys_2His_2 core (lower chemical potential), relative to a Cys_3His core.[33,49] Consistent with this observation, the development of *weak* electrophiles that do not affect Cys_2His_2 cores, yet disrupt NCp7, is a design strategy for NCp7 inhibitors.

The steric and electrostatic environments of the Sp1, NCp7, and GATA-1 zinc finger cores are illustrated in Figure 7, where the peptide backbones are color-coded according to the total electrostatic screening of the core per residue, $V_s^{residue}$. Except for the C-terminal NCp7 zinc finger, these proteins are representative of the average screening of Cys_2His_2, Cys_3His, and Cys_4 cores, based on V_s^∞ and ρ_s. In all cases there is positive (local) electrostatic screening of the core thiolates, primarily due to NH−S interactions. Moreover, Figure 7 emphasizes the enhanced local screening and backbone packing of the GATA-1 Cys_4 core, as well as an extended sphere of residues that substantially and positively screen its core ($Arg^{19,42,47}$, $Thr^{11,15}$), residues which are highly conserved in the GATA family of transcription factors. Compared to other fingers, the core of the C-terminal NCp7 finger is more exposed, particularly the Cys^{49} thiolate (S^{49}). This is more clearly shown

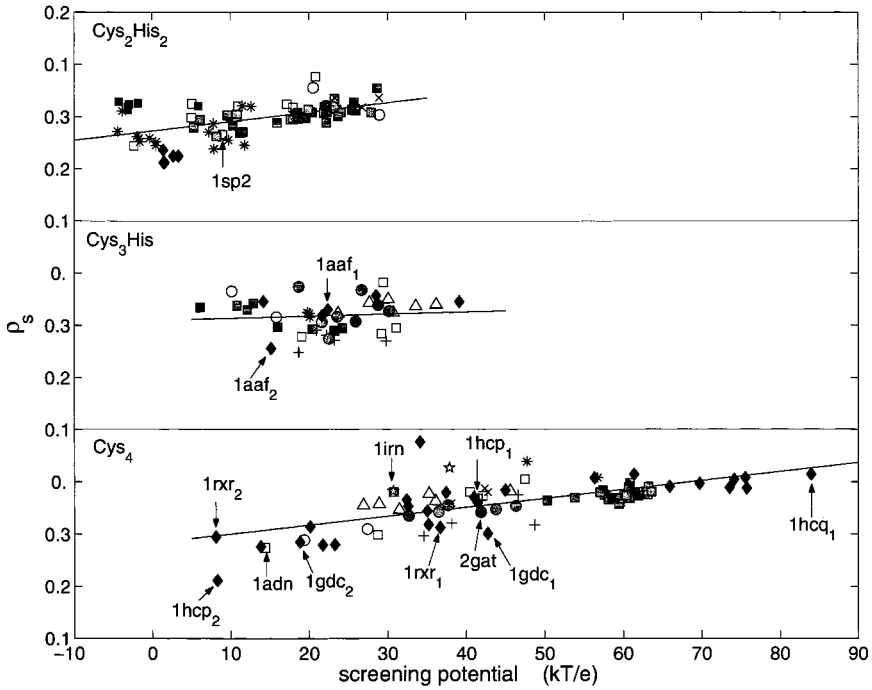

Figure 6. Diversity of protein screening of zinc finger cores, measured by (V_s^∞,ρ_s). Cores that are explicitly marked by their PDB labels are discussed in the text; PDB subscripts correspond to the sequence of multiple zinc fingers within a protein. (**Top panel**) Cys_2His_2 cores, protein families: yeast Adr1 TF (×), TFIIIA (*), GAGA TF (shaded ○), HIV-1 integrase (black ◇), enhancer binding protein (●), uridylyltransferase (black □), ZIF268 (shaded □), miscellaneous (□). (**Middle panel**) Cys_3His, protein families: adenovirus DNA-binding protein (△), CRP (+), RING-finger proteins (○), endonuclease I−PpoI (shaded □), protein kinase C (shaded ○), HIV-1 NCp7 (black ◇), transglycosylase (*), miscellaneous (□). (**Bottom panel**) Cys_4, protein families: ADH (shaded □), CRP (+), GATA TF (shaded ○), RING-finger proteins (○), adenylate kinase (×), adenovirus DNA-binding protein (△), rubredoxin (☆), polymerases (*), NR (black ◇), miscellaneous (□).

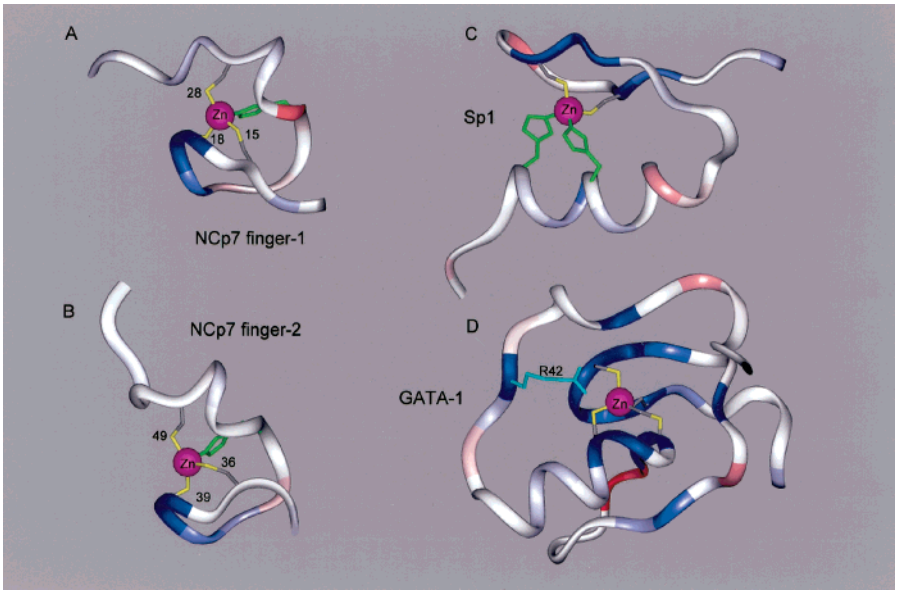

Figure 7. Comparison of the protein environments of NCp7 (**panel A**: N-terminal finger, **panel B**: C-terminal finger, PDB: 1aaf), Sp1 (**panel C**, PDB: 1sp2), and GATA-1 (**panel D**, PDB: 2gat) zinc finger cores. The backbone ribbons are color-coded by the contribution of each residue to the total electrostatic screening potential of the zinc finger cores; red (negative) to blue (positive), color saturation = ±2 kT/e.

in Figure 8, where electrostatic screening profiles of the individual core thiolates of NCp7 are compared, with $V_s(R)$ centered at each S position. Asymptotically, the electrostatic screening of S^{49} is 60 to 75% less than the other NCp7 thiolates. Experimentally, Cys^{49} is observed to be the most reactive cysteine of NCp7[21,22] and theoretically S^{49} is predicted to be the most nucleophilc site.[20]

The distributions $\{V_s^\infty,\rho_s\}^{motif}$ can be further used to identify other zinc fingers that may be chemically labile. Among Cys_2His_2 cores, V_s^∞ is nearly zero for the sterically exposed HIV integrase cores, and V_s^∞ is *negative* for the *E. coli* uridyltrasferase zinc fingers and the second TFIIIA fingers. Electrostatically, these cores are over one standard deviation less screened than the average of Cys_2His_2, $V_{sC_2H_2}^\infty$. Notably, there are no cases in which anionic core motifs are negatively screened (Figure 6). Among Cys_3His cores, the C-terminal finger of endonuclease I−PpoI and the RING finger domains of the promyelocytic leukemia and equine herpes virus TFs, are two

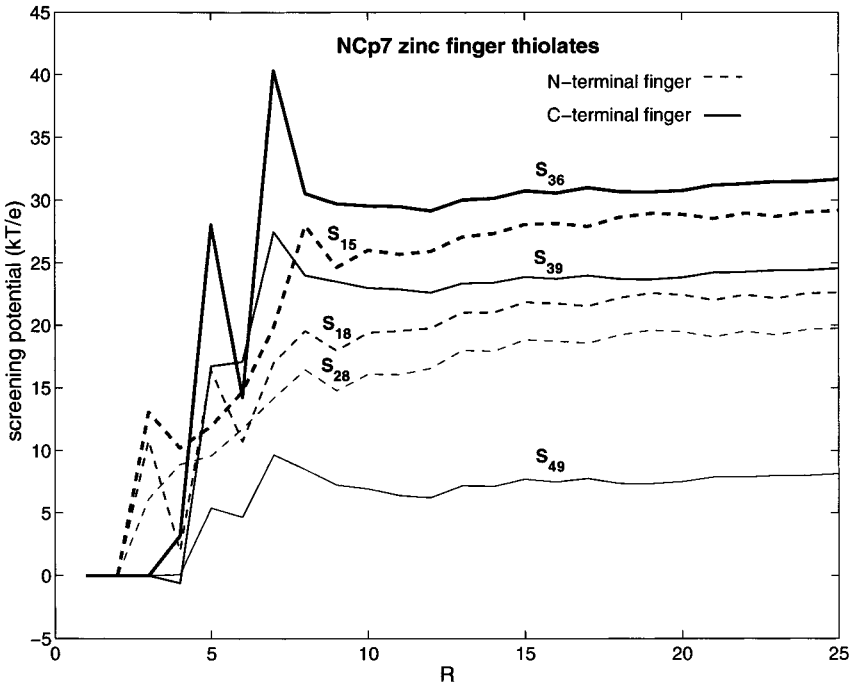

Figure 8. Radial profiles (Å) of the electrostatic screening potential (kT/e) of the Cys thiolates of NCp7, with the origin of $V_s(R)$ centered at the individual core S positions. Thiolates are labeled according to their Cys residue sequence.

standard deviations less screened than $V^{\infty}_{sC_3H}$. These Cys_3His cores and the C-terminal NCp7 core are predicted to be similarly labile. The $\{V^{\infty}_s,\rho_s\}^{Cys_4}$ distribution is distinguished by having the widest range of electrostatic and steric screening values. The V^{∞}_s range of Cys_4 is twice that of Cys_2His_2 and Cys_3His. The extrema of $\{V^{\infty}_s,\rho_s\}^{Cys_4}$ are spanned by the zinc fingers of nuclear receptor (NR) DNA-binding domains (DBDs). NRDBDs contain two distinct zinc finger domains; the minima of $\{V^{\infty}_s,\rho_s\}^{Cys_4}$ correspond to the C-terminal zinc finger cores of NR monomers and the maxima correspond to the N-terminal zinc cores of NR dimers.

In the case of NR *monomers* (Figure 6: 1gdc, 1hcp, 1rxr), the cores of the C-terminal zinc finger of the human retinoic acid receptor (hRR) and human estrogen receptor (hER) are over three standard deviations less electrostatically screened than $\bar{V}^{\infty}_{sC_4}$. Sterically, the C-terminal hER core is the least screened of Cys_4 cores, over four standard deviations less than the Cys_4 average of ρ_s. In contrast, the protein screening of the N-terminal zinc finger cores of hRR and hER is near the average of $\{V^{\infty}_s,\rho_s\}^{Cys_4}$. In the case of DNA-bound NR *dimers*, the protein screening of both N- and C-terminal cores is systematically increased, relative to NR monomers. For example, the C-terminal cores of hER and hRR dimers are shifted to the average of $\{V^{\infty}_s,\rho_s\}^{Cys_4}$, while the N-terminal cores are the most protected of this study (Figure 6: 1hcq). Thus, the C-terminal NR finger is predicted to be stabilized by NR dimerization.

The C-terminal zinc finger of monomeric NRs are predicted to be especially labile since they contain a Cys_4 core. As a point of reference, the reactive Cys_4 core of Ada[30] (Figure 6: 1adn) is located in the vicinity of the C-terminal zinc cores of hER, hRR, and glucocorticoid receptor. The C-terminal zinc finger of the hER DBD is predicted to be the most labile in this study (Figure 6: $1hcp_2$). Figure 9 illustrates the contrasting electrostatic and steric environments of the hER DBD Cys_4 cores, with the peptide backbone color-coded by $V^{residue}_s$. Notably, Arg^{63} is "sandwiched" between the two Cys_4 cores, forming a salt-bridge with the core of the N-terminal finger. Sequence alignments indicate that Arg^{63} is absolutely conserved across the NR family (data not shown). Figure 10 quantitatively compares the electrostatic screening of hER DBD thiolates, revealing a triad of Cys thiolates in the C-terminal finger (S^{49}, S^{59}, S^{62}) that are predicted to be particularly vulnerable to oxidation, S^{49} being the most labile. Electrostatically and sterically the C-terminal hER zinc core is 500 and 60% less screened than the N-terminal, respectively.

Experiments have demonstrated the two NRDBD zinc finger domains are structurally and functionally nonequivalent.[53−56] The N-terminal finger contains a DNA-recognition helix that interacts with the DNA major groove, while key components of the dimer interface are localized to the C-terminal finger.[53,54] NRDBDs are monomeric in solution, but dimerize in binding cooperatively and specifically to their DNA-response elements. As a monomer, the structure of the C-terminal finger is disordered, relative to the N-terminal, and on formation of the dimer the structure of the C-terminal finger is stabilized.[53,55] In addition, studies have found that hRR DNA-binding can be regulated by a variety of oxidants,[25,57] and that MT can reversibly exchange Zn^{2+} with hER, thus regulating its DNA binding.[26] The MT/thionein couple also effects DNA binding of Sp1[17] and TFIIIA,[17,58] and other experiments indicate that MT plays an important role in the cellular transduction of Zn^{2+}.[28,29,59] Consistent with these findings and the results of this study, we propose that redox regulation of NR DNA binding is modulated by the labile C-terminal finger. In this view, the C-terminal finger functions as a zinc-regulatory "switch", its Zn^{2+}-bound concentration controlled by the cellular redox state

(53) Schwabe, J.; Chapman, L.; Finch, J.; Rhodes, D.; Neuhaus, D. *Structure* **1993**, *1*, 187−204.
(54) Schwabe, J.; Chapman, L.; Rhodes, D. *Structure* **1995**, *3*, 201−213.
(55) Holmbeck, S.; Dyson, J.; Wright, P. *J. Mol. Biol.* **1998**, *284*, 533−539.
(56) Wikstrom, A.; Berglund, H.; Hambraeus, C.; van den Berg, S.; Hard, T.; *J. Mol. Biol.* **1999**, *289*, 963−979.
(57) Kroncke, K.-D.; Carlberg, C. *FASEB J.* **2000**, *14*, 166−173.
(58) Zeng, J.; Heuchel, R.; Schaffner, W.; Kagi, J. *FEBS Lett.* **1991**, *279*, 310−312.
(59) Jacob, C.; Maret, W.; Vallee, B. L. *Proc. Nat. Acad. Sci. U.S.A.* **1998**, *95*, 3489−3494.

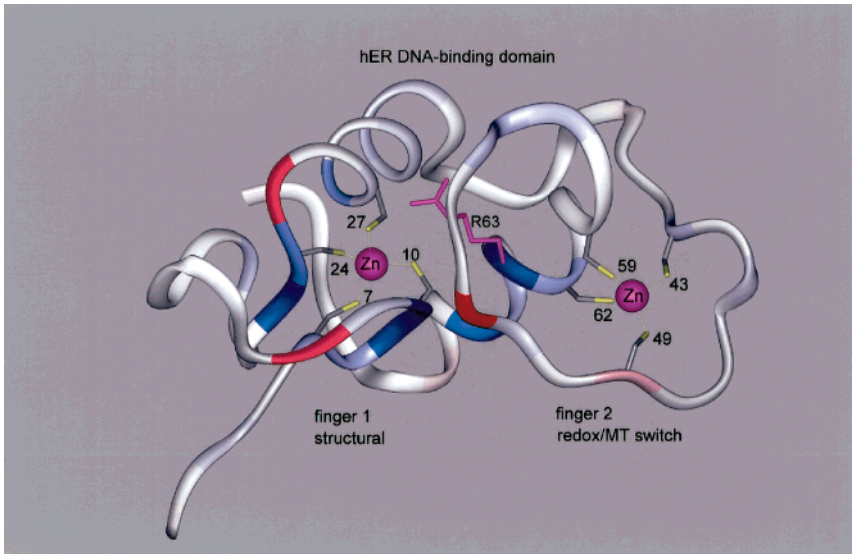

Figure 9. The zinc finger domains of the human estrogen receptor DNA-binding domain (finger 1 = N-terminal, finger 2 = C-terminal, PDB: 1hcp). The backbone ribbon is color-coded by the contribution of each residue to the total electrostatic screening potential of the zinc finger cores; red (negative) to blue (positive), color saturation = ±2 kT/e.

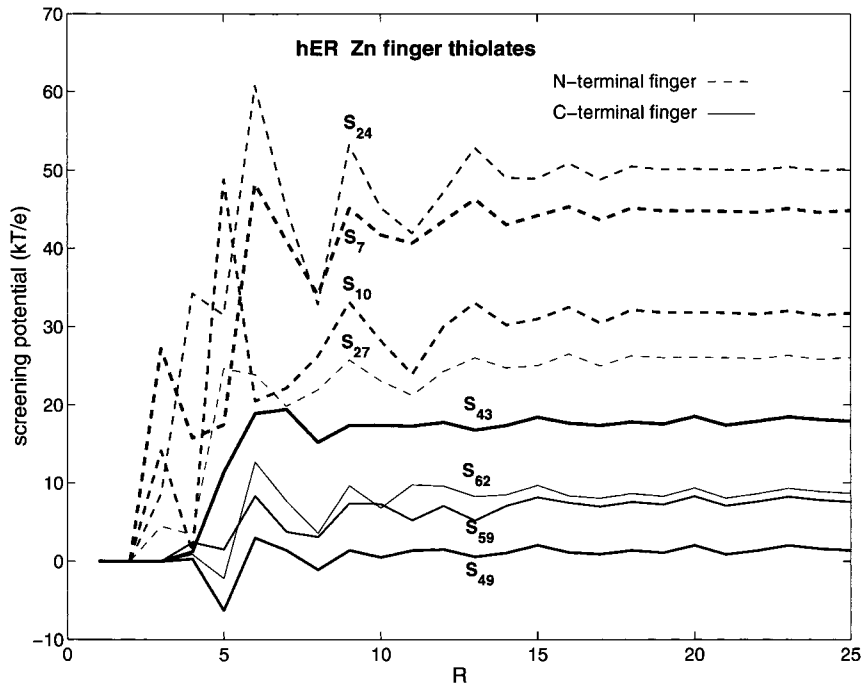

Figure 10. Radial profiles (Å) of the electrostatic screening potential of the core thiolates of the human estrogen receptor DBD. Thiolates are labeled by their Cys residue sequence.

or MT/thionein couple, which in turn regulates formation of the DBD dimer and hence DNA binding. Very recently, this hypothesis has been corroborated by Whittal and co-workers,[60] who demonstrated that preferential oxidation of the hER C-terminal zinc finger prevented hER dimerization and therefore DNA binding.

Conclusions

Using a structural database of zinc finger proteins, we have evaluated the chemical stability of zinc fingers on the premise that steric and electrostatic screening of their zinc cores confers resistance to oxidation. In support of this supposition, we find a large concomitant increase in the steric and electrostatic screening of increasingly anionic (reactive) core motifs, suggesting *structural* zinc fingers have evolved such that their potentially reactive core thiolates are protected by the protein, especially zinc fingers that utilize anionic cores. The predominant structural mechanism by which zinc finger cores are screened is through networks of backbone:core NH−S hydrogen bonds, which also stabilize the local protein fold of zinc fingers.[36,37] This finding is consistent with the notion of a secondary NH−S "coordination shell" in the region of zinc finger cores.[38] Although only the electrostatic component of NH−S interactions has been evaluated, the results suggest charge-transfer and polarization components may make important contributions to the oxidative stability of zinc cores. DFT and ab initio electronic calculations indicate NH−S interactions can substantially stabilize core thiolates; an optimized amide:

(60) Whittal, R.; Benz, C.; Scott, G.; Semyonov, J.; Bulingame, A.; Baldwin, M.; *Biochemistry* **2000**, *39*, 8406−8417.

Cys hydrogen bond lowers the Cys pK_a by −14 units.[49] Together, these findings support other studies[39−42] that indicate Cys_4 and Cys_3His cores are anionic under physiological conditions, but appear to contradict a recent mass spectrometric study[43] that indicates protonated (charge-neutral) Cys_4 and Cys_3His cores. Our analysis does not preclude the possibility of protonated cores, particularly unscreened cores which may be prone to *transient* protonation, thus providing a rationale for conflicting observations. Further studies are needed to resolve this issue. We also find that enhanced networks of core NH−S hydrogen bonds, associated with anionic cores, may induce increased backbone:backbone NH−O interactions. Furthermore, a number of ion pair [Arg,Lys]:[Cys_4,Cys_3His] interactions are noted that are highly conserved. Thus, we conclude zinc finger cores are not readily mutatable, since *each core motif determines specific protein:core interactions that contribute to unique zinc finger topologies*. Comprehensive mapping of protein:core interactions to residues may be instructive in interpreting sequence-conservation of zinc finger domains and have applications to the design of zinc fingers that are resistant to oxidation.[61,62]

By ranking zinc fingers according to the protein screening of their cores, potentially labile and stable zinc fingers are delineated. Consistent with experiment, the reactive Ada Cys_4 core[30] is identified as labile, as well as the C-terminal Cys_3His core of NCp7.[21−23] Lack of protein screening of anionic cores may be an indicator of zinc fingers with redox or MT regulated functionality, as found in the case of the C-terminal Cys_4 core of NR DNA-binding domains. The C-terminal zinc finger of hER is predicted to be the most labile in this study, containing a triad of reactive thiolates: S^{49}, S^{59}, S^{62}. Since the threshold of reactivity of the C-terminal hER core is predicted to be the lowest, it may be possible to selectively target the hER DBD with *weak* electrophiles, while not affecting other cellular zinc fingers. Thus, we propose that the transcriptional function of hER may be inhibited by *weak* electrophilic agents that modify its C-terminal zinc finger core, preempting formation of the hER dimer, thereby inhibiting hER DNA binding. This may provide an alternative strategy to current antihormone therapies, directed toward intervention of hER-mediated cancer growth.[63] Such a strategy is supported by the recent findings of Whittal and coworkers,[60] where hER DNA binding was inhibited through chemical modification of its labile C-terminal zinc finger.

Acknowledgment. This research was sponsored by the National Cancer Institute under Contract No. NO1-CO-56000. This publication does not reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The authors thank the staff of the Advanced Biomedical Computing Center at Frederick for resources and technical support. A.T.M. thanks L. G. Pedersen and A. A. Rabow for helpful comments and discussions.

JA0011616

(61) Walkup, G. K.; Imperiali, B. *J. Am. Chem. Soc.* **1997**, *119*, 3443−3450.

(62) Benson, D. E.; Wisz, M. S.; Hellinga, H. W. *Curr. Opin. Biotechnol.* **1998**, *9*, 370−376.

(63) Mori, H., Ed. *Oncology* **1998**, *54*, Suppl. 1, 1−59.