Quantum Descriptors for Biological Macromolecules from Linear-Scaling Electronic Structure Methods

Jana Khandogin and Darrin M. York[*]
Department of Chemistry, University of Minnesota, Minneapolis, Minnesota

ABSTRACT The characterization of electrostatic and chemical properties at the surface of biological macromolecules is of interest in elucidating the fundamental biological structure-function relationships as well as in problems of rational drug design. This paper presents a set of macromolecular quantum descriptors for the characterization of biological macromolecules in solution that can be obtained with modest computational cost from linear-scaling semi-empirical quantum/solvation methods. The descriptors discussed include: solvent-polarized electrostatic surface potential maps, equilibrated atomic charges, Fukui reactivity indices, approximate local hardness maps, and relative proton potentials. These properties are applied to study the conformational dependence of the electrostatic surface potential of the solvated phosphate-binding protein mutant (T141D), the regioselectivity of the zinc finger domains of HIV-1 nucleocapsid (NC) protein, and the order of pKa values of acidic residues in turkey ovomucoid third domain (OMTKY3) and of the zinc-binding residues in the carboxyl terminal zinc finger of NC. In all cases, insight beyond that obtainable from purely classical models is gained and can be used to rationalize the experimental observations. The macromolecular quantum descriptors presented here greatly extend the arsenal of tools for macromolecular characterization and offer promise in applications to modern structure-based drug design. Proteins 2004;56:724–737. © 2004 Wiley-Liss, Inc.

Key words: chemical reactivity index; pK_α descriptor; electrostatic potential; nucleocapsid protein; structure-based drug design

INTRODUCTION

Protein function is often modulated by binding to other molecules (or *ligands*) such as metal ions, small molecules, DNAs, and RNAs. The strength and specificity of the protein-ligand binding are governed by steric factors, such as size and shape, physicochemical interactions, such as hydrophobicity, hydrogen bonding, electrostatic and van der Waals forces, and chemical bonding. Over the past decade, structure-based drug design has evolved into a promising tool in medicinal chemistry,[1,2] fueled in part by extraordinary advances in three-dimensional protein structure determinations.[3]. An initial step in the structure-based drug design cycle involves obtaining a detailed knowledge of the geometric and physicochemical features of the protein receptor-binding site and of the principles of ligand-binding affinity and specificity.[2,4] Over the past decade, significant progress has been made in the development of computational methods for identification and analysis of ligand-binding sites.[5,6] However, the molecular features that give rise to selectivity remain elusive.[2]

Mapping the physicochemical properties of biological macromolecules onto their molecular surfaces offers a convenient means for visual inspection of the electrostatic and chemical characteristics of ligand-binding sites. Electrostatic surface potential (ESP) maps are routinely applied in structural biology and drug-design communities to display the electrostatic potential patterns of solvated proteins and nucleic acids and to aid qualitative prediction of ligand-binding sites based on the principle of electrostatic complementarity.[5,7,8] A common method to generate these maps is based on the finite-difference solution to the Poisson-Boltzmann equation (FDPB), in linearized or nonlinear form,[9–12] using a set of static atomic point charges and radii to model the macromolecular charge distribution and dielectric cavity, respectively.[13] The atomic charges and radii are typically parameterized against the experimentally measured solvation energies of small molecules[14] or else adopted from a molecular modeling force field. These methods have been applied with some success; however, they are limited in that they do not account for explicit electronic and solvent-induced polarization or the contribution from high-order atomic multipoles.

An alternative approach is to perform fully quantum mechanical calculations where the electron density is allowed to explicitly polarize and to adapt variationally to the solvated macromolecular environment. An advantage of quantum mechanical methods lies in their ability to provide, *in principle*, an accurate description of electronic response properties that are not explicitly included in conventional molecular mechanical models. For example, chemical reactivity plays an important role in the small-molecule binding to the zinc finger regions in HIV-1

nucleocapsid protein (NC)[15] and cannot be obtained from the classical molecular modeling methods. Molecular electrostatic potentials and local ionization and polarization energies obtained from conventional quantum electronic structure calculations have proved useful for the characterization of noncovalent interactions and reactivities of small biological molecules.[16,17] The ever lowering cost of high-speed computational resources and the advances in efficient linear-scaling algorithms[18−24] have made possible the application of quantum methods, mainly at the semi-empirical level at the present time, to solvated macromolecules containing tens of thousands atoms.[25−27] Recently, linear-scaling semi-empirical methods combined with conductor-like screening model (COSMO)[28−30] have been applied to study solvent effects on the electrostatic surface potential of nucleic acids,[27] and the electrostatic potential as well as local chemical reactivity of two zinc finger regions in NC.[31]

Protein stability and function are dependent on environmental pH.[32] The knowledge of the pKa's of ionizable side-chains in a protein is important in structure-based drug design, as ligand binding can be coupled to the protein protonation state. A rigorous determination of pKa value of a titratable residue involves formally 2^N solvation free energy calculations, where 2 accounts for the neutral and charged states and N is the total number of protonation sites. In this report, the electrostatic potential at a titratable proton site (proton potential) is proposed as a molecular descriptor for qualitative prediction of the order of pKa's of the weakly-coupled multiple ionizable groups in a protein. The electrostatic surface potential[33] and the proton potential[34] have been previously suggested as hydrogen bond descriptors for small molecules. The advantage of using proton potential as a pKa descriptor lies in the fact that it can be obtained without additional computational cost as a by-product of a linear-scaling electronic structure calculation. This approach is particularly suitable for the identification of abnormally shifted pKa's, which can be found often at or near ligand-binding sites.[35] For instance, the most reactive thiolate in NC's C-terminal zinc finger was linked to its unusually high pKa value.[36]

Chemical hardness and softness are well-known concepts in inorganic and acid/base chemistry, and are frequently used to rationalize the reactivity of small molecules and ions (see, for example, a more recent work[37]). Two related quantities for measuring regioselectivities are the local softness (or the related Fukui index) and local hardness, and have been recently used to probe the gas-phase chemical reactivities of small organic molecules.[38−41] Applications to molecules of biological interest are emerging. Fukui indices were used to probe the reactive sites in DNA bases[42] and protein zinc fingers.[43] However, these reactivity indices have not been applied to macromolecular systems due to limitations in the scalability of conventional electronic structure methods that have only recently been overcome with advances in linear-scaling quantum methods.[18−24]

The purpose of this report is twofold: (1) to introduce a set of *macromolecular* quantum descriptors from linear-scaling electronic structure and solvation methods, and (2) to demonstrate that these descriptors provide new chemical insight into biological systems that can not be obtained by classical methods. The macromolecular properties introduced in this work include solvent-polarized electrostatic potential maps, equilibrated atomic charges, and proton potentials, Fukui indices, and approximate local hardness maps. These descriptors are distinctive from those applied previously for biological molecules in that they simultaneously take into account the macromolecular and solvent environment. The report is organized as follows. Methods briefly describes the methods and computational procedure. Results and Discussion provides validation and convergence of the new methods, and applies them to several biologically important proteins that have been well studied by both experimental and theoretical means. First, the solvent-polarized ESP maps of the phosphate-binding protein mutant T141D and the NC protein are compared to those obtained from the FDPB method with different sets of atomic charges. Second, solvent-polarized atomic charges, Fukui indices, and approximate local hardness maps are obtained for NC and NC-RNA complex and compared to experimental results to explain the regioselective reactivity of the zinc-binding domains. Finally, relative proton potentials at the acidic sites in ovomucoid turkey third domain (OMTKY3) and the C-terminal zinc-binding residues of NC are correlated with the experimental pKa's. The Conclusion draws together the main conclusions and projects new directions for the application of linear-scaling electronic structure methods in understanding biological function and in structure-based drug discovery. A concise discussion of the theoretical background and essential technical details of the implementation are provided in the Appendix.

METHODS
Quantum Descriptors

This subsection summarizes the quantum descriptors for biological macromolecules applied in this work. A detailed discussion of the theoretical background and derivations can be found in the Appendix.

Solvent-polarized electrostatic potential and atomic charges

In the present work, the linear-scaling semi-empirical quantum method is coupled to a smooth[30,44] linear-scaling[29] conductor-like screening model (COSMO).[28] The model uses a variational condition on the electrostatic energy to solve approximately the Poisson equation for the solvent reaction field potential.[30] The electronic charge density of the quantum mechanical solute is polarized by the solvent reaction field, and subsequent (approximate) solution of the Poisson equation provides a *solvent-polarized electrostatic potential*. The explicit electronic polarization of the solute, which is neglected in conventional macromolecular applications using nonpolarizable force field models, is made possible via linear-scaling electronic structure theory.

Solvent-polarized electrostatic potential can be mapped onto molecular surfaces for graphic display and mapped

onto atomic sites for quantitative studies. Two particularly useful quantities are *atomic charges* and *proton potentials*. The former provides information about the local electronic and solvent environment around individual atoms in a macromolecule. The latter provides information about the relative acidity of titratable protons. In this work, atomic charges are calculated using the CM2 charge model.[45] The CM2 charges have been used to describe the electrostatic interactions in QM/MM applications,[46] to predict solvation polarization energies with Generalized-Born (GB)[45,47] and Poisson-Boltzmann (PB) solvation models,[48] and to examine charge fluctuations in solvated DNA and RNA systems.[49]

Local hardness, softness, and Fukui functions

Density-functional theory provides a framework for the derivation of a host of response functions from the electron density. For example, the Mulliken *electronegativity* defined as $\chi = 1/2(IP + EA)$, where IP and EA are the ionization potential and electron affinity, respectively, can be viewed as a finite difference approximation to minus the *chemical potential* $\mu = \partial E/\partial N$ of density functional theory. Similarly, the Pearson's *chemical hardness* scale[50] is defined to be $\eta = IP - EA$, which is a finite difference approximation to the DFT derivative $\eta = \partial^2 E/\partial N^2$. The *chemical softness* is simply the inverse of the chemical hardness.

The chemical hardness, softness, and electronegativity are inter-related and of significance for understanding the chemical reactivity of a molecule. These quantities can be formulated as integrals of spatial functions known as the *local hardness* and *local softness* for characterizing the reactivity in a spatial region of the molecule. The local softness divided by the global softness is called the Fukui function, and has relation to the frontier orbital theory. Local hardness, local softness/Fukui function can be formulated as derivatives of the chemical potential and electron density. Analogous to atomic charges, Fukui functions can be mapped onto atoms resulting in Fukui indices in the same manner as the electron density can be partitioned to obtain atomic charges. Together, the local softness/Fukui function and local hardness provide insight into regioselective reactivities of macromolecules.

Structure Preparation and Computational Protocols

All macromolecular descriptors were implemented into the linear-scaling Divide-and-Conquer (D&C) semi-empirical program[29,51] with solvent effects treated by the linear-scaling COSMO solvation method.[28,29] The CM2 electron density partition scheme[45] was implemented here to derive atomic charges and Fukui indices. The proton potential implemented here is a sum of three contributions: the potential due to all solute nuclei but one proton, the potential due to all electrons, and the potential due to the solvent reaction field. Note that the component of the solute potential contains a constant term due to the total kinetic energy. The latter cannot be separated from the one-electron nuclear-electron attraction integral in the

NDDO (Neglect of Diatomic Differential Overlap) approximations that are the basis for the AM1, PM3, and MNDO methods. Therefore, the proton potential is referred to as the relative proton potential throughout the subsequent discussions.

The calculations of the atomic charges, Fukui indices, and electrostatic potential of NC were based on 25 NMR structures of the NC-SL3 RNA complex (PDB ID: 1A1T,[52] about 1,500 atoms including hydrogens). The hydrogen positions of all 25 structures were initially optimized until the default convergence using the CHARMM program[53,54] with the CHARMM22 force field for proteins[55] and CHARMM27 force field for nucleic acids.[56] Electrostatic interactions were treated with a distance-dependent dielectric function ($\epsilon = r$), and no cutoffs were used for non-bonded interactions. The parameters for Zn^{2+} were taken from Stote and Karplus.[57] The hydrogen optimized structures were relaxed by 100 steps of steepest descent on quantum surface using the linear-scaling semi-empirical program with the PM3 Hamiltonian, which was previously shown to give reasonable geometries for Zn^{2+}-complexes.[26] In the linear-scaling energy minimizations, the subsystem was defined to be a single amino acid or nucleotide and a buffer/matrix cutoff of 6/7 Å was used.[29,51] The SCF convergence criterion was set at 10^{-5} kcal/mol/ atom. The COSMO calculations were performed with a set of radii specifically parameterized for biomolecules.[29] The structures with the lowest heat of formation ($< -10,000$ kcal/mol) after steepest descent quantum energy minimizations (19 structures in total) were chosen for the subsequent single-point calculations for the macromolecular descriptors using a buffer/matrix cutoff of 8/9 Å. Here, RNA was excluded from the calculation. The solvent-polarized potential and approximate local hardness were computed on a $64 \times 64 \times 64$ grid and mapped onto a molecular surface and graphically displayed with the program GRASP.[58] The molecular surface used was defined by the overlapping atomic spheres with the same set of radii as used in the solvation calculations. The ESP maps of NC in the SL3-RNA bound conformation shown in this report were based on entry no. 1 of the NMR structure ensemble. The approximate local hardness of the NC-SL2 RNA complex was calculated based on entry no. 1 of the NMR structure ensemble (PDB ID: 1F6U[59]), using the same energy minimization protocol as described above.

The calculation of the electrostatic potential on the phosphate-binding protein mutant T141D was based on the ligand-free (PDB ID: 1OIB[60]) and ligand-bound (PDB ID: 1IXG[61]) crystal structures. Hydrogen positions were added and optimized until the default convergence using the CHARMM program. The hydrogen-optimized structures of both the free and ligand-bound forms of T141D were subjected to the linear-scaling quantum calculations using the previously mentioned cutoff and the AM1 Hamiltonian, which is known to better reproduce hydrogen bonded complexes.[62] Here, the phosphate was not included in the single-point quantum calculations.

The calculation of the relative proton potentials on the carboxylic groups in OMTKY3 was based on entry no. 1 of

TABLE I. Convergence of the RMSD of the Electrostatic Potential, Fukui Index, Approximate Local Hardness, and Relative Proton Potential With Respect to the Buffer/Matrix Cutoff

	4/5	6/7	8/9	10/11	12/13
$\eta \backslash \phi$					
4/5	—	1.101	1.112	1.118	**1.118**
6/7	4.361×10^{-1}	—	1.551×10^{-2}	2.005×10^{-2}	$\mathbf{2.009 \times 10^{-2}}$
8/9	1.394×10^{-1}	5.662×10^{-1}	—	5.493×10^{-3}	$\mathbf{5.534 \times 10^{-3}}$
10/11	2.044×10^{-1}	6.281×10^{-1}	7.264×10^{-2}	—	$\mathbf{4.163 \times 10^{-5}}$
12/13	$\mathbf{2.199 \times 10^{-1}}$	$\mathbf{6.448 \times 10^{-1}}$	$\mathbf{8.639 \times 10^{-2}}$	$\mathbf{1.705 \times 10^{-2}}$	—
$f_i \backslash q_i$					
4/5	—	4.147×10^{-3}	3.980×10^{-3}	3.982×10^{-3}	$\mathbf{3.982 \times 10^{-3}}$
6/7	8.312×10^{-3}	—	1.312×10^{-3}	1.320×10^{-3}	$\mathbf{1.321 \times 10^{-3}}$
8/9	1.974×10^{-2}	1.827×10^{-2}	—	3.526×10^{-5}	$\mathbf{3.612 \times 10^{-5}}$
10/11	2.024×10^{-2}	1.903×10^{-2}	2.992×10^{-3}	—	$\mathbf{2.970 \times 10^{-6}}$
12/13	$\mathbf{2.055 \times 10^{-2}}$	$\mathbf{1.937 \times 10^{-2}}$	$\mathbf{2.977 \times 10^{-3}}$	$\mathbf{3.756 \times 10^{-4}}$	—
$\backslash \phi_i^H$					
4/5	—	2.542×10^{-1}	2.542×10^{-1}	2.541×10^{-1}	$\mathbf{2.541 \times 10^{-1}}$
6/7	—	—	4.094×10^{-3}	4.188×10^{-3}	$\mathbf{4.187 \times 10^{-3}}$
8/9	—	—	—	2.920×10^{-4}	$\mathbf{2.932 \times 10^{-4}}$
10/11	—	—	—	—	$\mathbf{1.021 \times 10^{-5}}$
12/13	—	—	—	—	—

†RMSD's between the macromolecular descriptor vectors calculated with different buffer/matrix cutoff schemes (shown as row and column headers in Å) in the linear-scaling D&C method. The RMSD of a vector Δx is defined as $\langle \Delta x^2 \rangle$, where here Δx is the difference between two descriptor vectors. The upper triangles (the numbers above the fields with dash signs) present the RMSD convergence for the electrostatic potential $\phi(r_l)$, atomic charges q_i, and relative proton potential $\phi_i^H (r_i)$, while the lower triangles present the convergence for the local hardness $\eta^- (r_l)$ and Fukui index f_i^-. Here, the subscript l and i refer to a grid point and an atom, respectively. Rows and columns that compare to the highest level cutoff scheme (12/13 Å) are shown in bold and are reasonable indicators of the convergence level of the lower order schemes. All quantities are in atomic units except for $\phi(r_l)$ and $\eta^- (r_l)$, which are given in units of $k_B T/e$ (k_B = Boltzmann constant; e = electron charge; T = 298 K).

the NMR structure ensemble (PDB ID: 1OMU[63]). First, the carboxylic acid oxygen of each acidic group was determined by comparing the minimized energies of two possible protonation states with all other amino acid residues in their standard protonation states at neutral pH. Then, the whole protein, with protonated acidic and basic groups, was subject to an energy minimization with all coordinates except for the five acidic residues held fixed. The converged structure was used for the subsequent linear-scaling semiempirical calculation with the AM1 Hamiltonian. The calculation of the relative proton potentials on the NC protein in the SL3-RNA bound conformation was based on entry no. 1 of the NMR structure ensemble (PDB ID: 1A1T,[52] RNA excluded). First, Zn^{2+} was removed from the C-terminal zinc finger domain. Then, protons were added to the thiolates and all hydrogen positions were optimized until the default convergence using the CHARMM program and CHARMM22 force field.

RESULTS AND DISCUSSION

In this section, the stability and convergence of the linear-scaling quantum descriptors are demonstrated and applications to several important biological systems are shown. This is among the first times[31] that such electrostatic and chemical descriptors have been applied with the linear-scaling quantum methods to biological macromolecules in solution. Agreements with experimental results suggest that they provide insight beyond that obtainable from purely classical models and offer considerable promise toward the development of new-generation techniques for macromolecular characterization and rational drug design.

Convergence of Calculated Macromolecular Descriptors With the Linear-Scaling Method

The convergence behavior of the calculated atomic charge, Fukui index, electrostatic potential, approximate local hardness, and relative proton potential with respect to the buffer/matrix cutoff is presented in Table I as the root-mean-square deviation (RMSD) between the descriptor vectors calculated with different levels of cutoffs. The upper triangles in Table I show the RMSD's of the electrostatic potential, atomic charge, and relative proton potential while the lower triangles for the approximate local hardness and Fukui index. The convergence of the RMSD's for the atomic charge and electrostatic potential is one or two orders of magnitude faster than that for the Fukui index and approximate local hardness, respectively. With the buffer/matrix cutoff of 8/9 Å, the error is reduced to the order of 10^{-3} and 10^{-2} $k_B T/e$ for the electrostatic potential and local hardness, 10^{-5} and 10^{-3} e for the atomic charge and Fukui index, respectively, and 10^{-4} V for the relative proton potential. These errors are well within the acceptable range for practical applications. Therefore, a 8/9 Å buffer/matrix cutoff is used in the subsequent applications, consistent with previous work.[27,49,64]
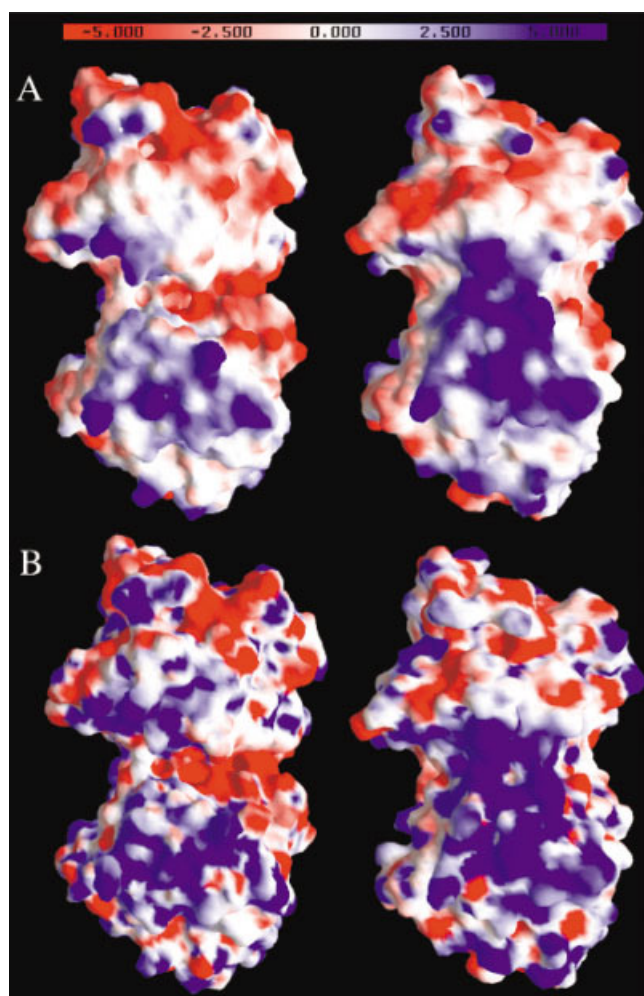
Fig. 1. Electrostatic surface potential (ESP) maps of the free and ligand-bound forms of the phosphate-binding protein mutant T141D obtained from the AM1/COSMO (**A**) and PARSE/PB (**B**) methods. Shown on the left are the ESP maps of the free structure (PDB ID: 1OIB[60]); shown on the right are those of the ligand-bound structure (PDB ID: 1IXG[61]). Proteins are rotated such that the phosphate binding cleft is facing toward the reader and domain I is at the top and domain II is at the bottom. The electrostatic potential is given in units of $k_B T/e$ (k_B = Boltzmann constant, T = 298 K, and e = charge of an electron).

Electrostatic Surface Potential
Application to the phosphate-binding protein

The calculation of electrostatic potentials to aid the prediction of ligand or metal binding sites in proteins and nucleic acids has demonstrated considerable promise in recent years.[7,13] Frequently, the receptor macromolecule is assumed to retain a bound-state conformation,[7] which may differ significantly from its free form.[60] Consequently, it is important to examine how the electrostatic potential pattern of the receptor alters in response to changes that occur upon ligand binding.

Consider the ESP maps of the ligand-bound and free forms of the phosphate-binding protein mutant T141D, which have been studied previously using the FDPB method with the full charge set, referred to as the Full/PB map.[60] Here, the full charge set is defined as the charge
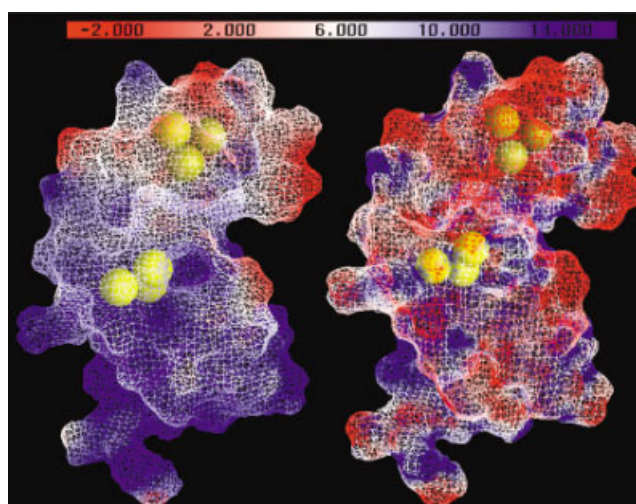

Fig. 2. Electrostatic surface potential maps of HIV-1 NC in the SL3-RNA bound conformation obtained from the PM3/COSMO (**left**) and PARSE/PB (**right**) methods. The protein is oriented such that the best view to the Zn^{2+}-coordinating thiolates (displayed as yellow spheres) can be obtained. The C-terminal zinc finger region is at the top and the N-terminal one is at the bottom. The electrostatic potential is given in units of $k_B T/e$.
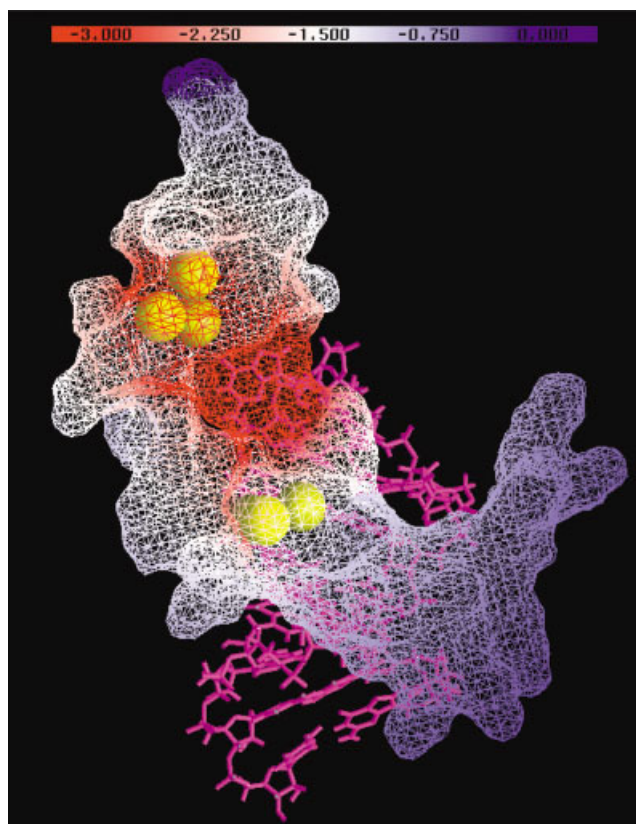

Fig. 4. The approximate local hardness map of NC bound to SL2-RNA. The protein is oriented such that the best view to the Zn^{2+}-coordinating thiolates (displayed as yellow spheres) can be obtained. The C-terminal zinc finger region is at the top and the N-terminal one is at the bottom. The approximate local hardness $\eta_F(r)^-$ is given in units of $k_B T/e$. Shown is the negative of this quantity in order to emphasize the relation between the electron density, which has a negative charge, and the finite-difference Fukui function from which the approximate local hardness is derived.

assignments to the ionizable residues as whole entities. This is a default charge set for the DelPhi program,[9] which has been widely used in the structural biology and molecular modeling community to obtain information about the solvated electrostatic potential and related properties in biological macromolecules. The objective here is to compare the ESP patterns of the free and ligand-bound forms of T141D generated with the linear-scaling AM1/COSMO method, referred to as the AM1/COSMO map, and the FDPB method with PARSE charges and radii, referred to as the PARSE/PB map. PARSE charges were parameterized specifically for the FDPB method using a database of experimental solvation energies of small molecules.[14] Figure 1(A) displays the AM1/COSMO maps for the ligand-free (shown on the left) and the ligand-bound structures (right). Analogously, Figure 1(B) displays the PARSE/PB maps based on the free (left) and ligand-bound (right) structures.

Consider the AM1/COSMO maps in Figure 1(A). In the unliganded structure (left), the two domains are far apart exposing a deep binding cleft containing residues that can make strong hydrogen-bond contacts with the phosphate upon binding. Except for Asp56, all these residues occupy regions of positive electrostatic potential in solution, despite the negative potential in the nearby cleft region. The apex regions of both domain I (top) and domain II (bottom) display predominantly positive electrostatic potential. In the ligand-bound structure (right), the apex regions of the two domains are brought together burying the phosphate ligand and display considerable positive electrostatic potential, in contrast to those of the unbound form.

Compared with the AM1/COSMO maps, the PARSE/PB maps show no significant difference. However, the surface potential from the PARSE/PB method appears to be less continuous and is overall larger in magnitude. These differences can be traced back to the difference in the solute charge representation: static point charges are used for PARSE/PB maps and solvent-polarized electron density is used for AM1/COSMO maps. It is worthwhile to notice that both the AM1/COSMO and PARSE/PB maps of the unliganded structure are in sharp contrast to the Full/PB map.[60] The Full/PB map[60] shows an intensely negative potential everywhere at the binding cleft, on the basis of which the validity of the electrostatic complementarity was brought into question for the phosphate-protein binding.[60] In contrast, both the AM1/COSMO and PARSE/PB maps display positive potential on residues in the binding cleft. This emphasizes the dependence of the ESP pattern on the solute charge and cavity model for biological macromolecules. ESP maps generated with the PARSE parameters (PARSE/PB) agree reasonably with those obtained with the AM1/COSMO approach. The difference between the Full/PB and PARSE/PB maps and the closer agreement of the latter with the AM1/COSMO map are easily appreciated, since the PARSE charges are obtained specifically for more accurate PB solvation energies.

It is worthwhile to mention that there are different models for determination of the electrostatic potential in

TABLE II. Atomic Charges on the Zn^{2+}-Coordinating Atoms in NC of the SL3-RNA-Bound Conformation†

Atom	Residue	CM2 charge mean (RMSD)	Mulliken charge mean (RMSD)
Sγ	Cys15	$-0.512\ (5.63 \times 10^{-3})$	$-0.438\ (9.15 \times 10^{-3})$
Sγ	Cys18	$-0.459\ (2.24 \times 10^{-3})$	$-0.382\ (7.91 \times 10^{-3})$
N$\varepsilon2$	His23	$0.062\ (8.76 \times 10^{-3})$	$0.269\ (1.55 \times 10^{-2})$
Sγ	Cys28	$-0.415\ (7.07 \times 10^{-3})$	$-0.338\ (7.32 \times 10^{-3})$
Sγ	Cys36	$-0.478\ (4.94 \times 10^{-3})$	$-0.403\ (6.24 \times 10^{-3})$
Sγ	Cys39	$-0.476\ (8.14 \times 10^{-3})$	$-0.401\ (8.20 \times 10^{-3})$
N$\varepsilon2$	His44	$0.065\ (6.10 \times 10^{-3})$	$0.277\ (2.14 \times 10^{-2})$
Sγ	Cys49	$-0.408\ (1.39 \times 10^{-2})$	$-0.328\ (1.97 \times 10^{-2})$

†Atomic charges are given in units of e. RMSD's due to the conformational variation in the NMR ensemble structure are given in parentheses.

semi-empirical methods,[65,66] in addition to different semi-empirical Hamiltonians that continue to improve.[62,67] In this work, the potential calculated directly from the electron density (as an atomic multipole expansion) was employed throughout. It was observed that the ESP pattern is fairly robust with respect to the semi-empirical Hamiltonian, although the magnitude of the potential can vary to a greater degree (data not shown). A systematic study of the ESP patterns using different Hamiltonians and electrostatic potential models is beyond the scope of the present work.

Application to HIV-1 nucleocapsid protein

As a second example of comparison between the ESP patterns generated with the quantum and classical methods, both the PM3/COSMO (left) and PARSE/PB (right) maps of NC in the SL3-RNA bound conformation are displayed in Figure 2. In contrast to the PARSE/PB map, the PM3/COSMO map is able to distinguish the electrostatic environment of the two zinc finger regions. The C-terminal zinc finger region is clearly more electronegative than the N-terminal zinc finger. This is in agreement with previous theoretical[68] and experimental work[69−71] that suggest the C-terminal zinc finger to be more prone to electrophilic attack. A detailed analysis of the electrostatic features of NC can be found elsewhere.[31]

Solvent-polarized atomic charge

Table II shows the conformational averages and RMSD's of the atomic charges on the Zn^{2+}-coordinating atoms in NC in the SL3-RNA bound conformation calculated from the Mulliken and CM2 electron density partitioning schemes. Although the atomic charges resulting from the two schemes have slightly different magnitudes, the order among different atoms remains the same: among three Zn^{2+}-binding thiolates, the Cys15 thiolate (S15) from the N-terminal finger has the most negative charge and the Cys49 thiolate (S49) from the C-terminal zinc finger has the least negative charge. The destabilization of the negative charge on S49 may be attributed to the negatively charged neighbor Asp48, consistent with the less positive electrostatic surface potential around the C-terminal zinc
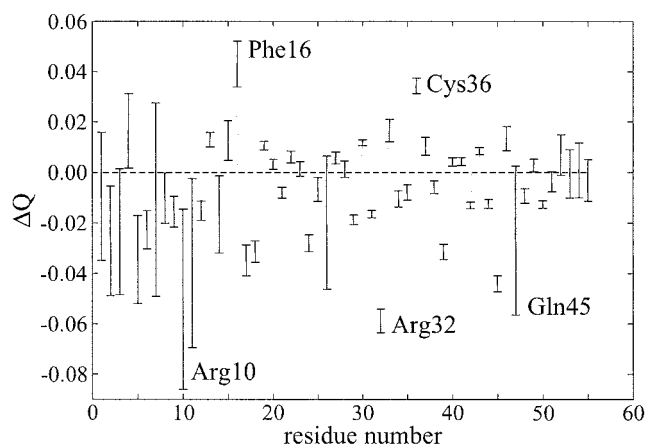
Fig. 3. RNA binding induced charge polarization on NC. $\Delta Q_i = Q_i(\text{bound}) - Q_i(\text{free})$, where $Q_i(\text{bound})$ and $Q_i(\text{free})$ represent the average CM2 charge of the ith residue over the NMR conformation ensemble in the RNA-bound and free protein, respectively. The "error bars" indicate the RMSD's due to the conformational variations.

finger region as compared to the N-terminal finger. Based on the electrostatic surface potential and atomic charge results, S49 is predicted to be electrostatically the least stable Zn^{2+}-binding thiolate, in agreement with the electrostatic screening study using the FDPB method.[68]. The RMSD's of the atomic charges due to the conformational variation within the NMR structure ensemble are mostly small (Table II), on the order of $10^{-3}\ e$. The greater RMSD on S49 is correlated with the pronounced conformational flexibility of the C-terminal region of the protein.[31]

Figure 3 shows the RNA-induced polarization of the residue charges of NC. Overall, the changes are relatively small, on the order of $10^{-2}\ e$. As expected, greater changes occur on residues that contain atoms capable of intermolecular hydrogen bonding (Fig. 3). The total CM2 charge of the protein is decreased by $0.448\ e$ upon binding (roughly the same with the Mulliken charge model), and reflects a net charge transfer from RNA to protein. Since the protein in the isolated form is positively charged and the RNA is negatively charged, a decrease in the net charges of both species minimizes the electrostatic repulsion and results in a stabilization of the complex, consistent with the density of states analysis discussed later. Charge transfer between the protein and solvent water has been previously studied using the linear-scaling semi-empirical calculation.[25]

Fukui Index and Approximate Local Hardness

Often, approximate Fukui functions for small molecules are calculated with a frozen orbital assumption and the finite difference approximation assuming a unit step size ($\Delta N = 1$, see Eq. 12). A density functional theory–based study[72] revealed that when orbital relaxation is allowed, a smaller step size, such as $\Delta N = 0.01$, is more reliable for prediction of the precise spatial reactivity preference in small molecules. The present work is restricted to the frozen orbital approach and in what follows the effect of the finite difference step size on the Fukui function is

examined. This is worthwhile since biological macromolecules contain a very large number of electrons and many nearly degenerate states close to the HOMO and LUMO orbitals. Preliminary tests showed that using a step size of 1 lead to somewhat arbitrary spatially specific index that was highly sensitive to the conformational variations exhibited by the NMR ensemble of structures. A more robust index was sought that was able to reveal the reactive "hot spots" spatially without necessarily pinpointing *a single very localized reactive region* that might be sensitive to conformational variations. Toward this end, different finite-difference step sizes were considered. Table III summarizes the results of the Fukui indices for NC in the SL3-RNA bound conformation using the step sizes $\Delta N = 1,5,10$. With $\Delta N = 1$, the largest average Fukui index is found on the S49 and the second largest on the C37 (Cγ atom of Trp37). The large RMSD's of the Fukui indices on the atoms close to the C-terminus are consistent with the greater variations in their atomic charges as discussed earlier. It is interesting to see that most of the large Fukui indices are located on the Zn^{2+}-coordinating thiolates. This is due to the fact that the high-lying occupied orbitals spatially localize on the Zn^{2+}-coordinating Cys residues, as will be discussed later. With $\Delta N = 5$ and 10, the maximal Fukui index remains on S49. Since the left Fukui function measures the reactivity of a spatial region in the molecule toward electrophilic attack, the Fukui indices reveal the Cys49 thiolate to be most reactive toward electrophilic attack, and provide rationale for the experimental finding that Cys49 is the initial attacking site for electrophilic reactions.[69–71] The Fukui index results (Table III) are also consistent with a previous density functional study using the small zinc finger model structures.[43] It has been reported in the literature that the relative order of Fukui indices somewhat depends on the electron density partition schemes.[73] In the case of semi-empirical methods, these variations are minimal for the systems studied here.

In our definition (Eq. 20), the approximate local hardness $\eta_F(\mathbf{r})^-$ is closely related to the Fukui function and embodies both the chemical reactivity and electrostatic potential. The interpretation of the local hardness is not as straightforward as that of the global hardness or the chemical hardness of a small molecule, which can be thought of as the resistance to a charge transfer,[74] since the definition of the local hardness is somewhat ambiguous (as discussed above). Previously, the approximate local hardness derived from the electron density $\eta_D(\mathbf{r})^-$ was found to have good correlation with the intermolecular reactivity sequence of some small organic molecules.[38,39] The approximate local hardness $\eta_F(\mathbf{r})^-$ has also been invoked to explain the regioselectivity in some Diels-Alder reactions.[41] In a recent study,[31] $\eta_F(\mathbf{r})^-$ projected onto the molecular surface was used to correlate to the most probable site for electrophilic attack in the NC-SL3 RNA complex. Figure 4 shows the $\eta_F(\mathbf{r})^-$ map of NC bound to HIV-1 SL2-RNA, where the C-terminal zinc finger region (red) indicates the concentration of large values of approximate local hardness, in agreement with the previous study

TABLE III. Fukui Indices for the NC of the SL3-RNA-Bound Conformation With Different Finite-Difference Step Sizes[a]

Atom name	Residue	$q_A - (\Delta N = 1)$ mean (RMSD)	$q_A - (\Delta N = 5)$ mean (RMSD)	$q_A - (\Delta N = 10)$ mean (RMSD)
Sγ	Cys15	$0.026\,(3.48 \times 10^{-2})$	$0.067\,(4.45 \times 10^{-2})$	$0.103\,(3.14 \times 10^{-2})$
Sγ	Cys18	$0.028\,(6.81 \times 10^{-2})$	$0.078\,(5.64 \times 10^{-2})$	$0.101\,(2.76 \times 10^{-2})$
Sγ	Cys28	$0.013\,(3.44 \times 10^{-2})$	$0.022\,(1.83 \times 10^{-2})$	$0.045\,(3.50 \times 10^{-2})$
Sγ	Cys36	$0.031\,(3.43 \times 10^{-2})$	$0.074\,(4.56 \times 10^{-2})$	$0.137\,(4.22 \times 10^{-2})$
Cγ	Trp37	$0.116\,(9.98 \times 10^{-2})$	$0.097\,(3.65 \times 10^{-2})$	$0.056\,(1.06 \times 10^{-2})$
Nε1	Trp37	$0.090\,(7.83 \times 10^{-2})$	$0.074\,(2.82 \times 10^{-2})$	$0.043\,(8.06 \times 10^{-3})$
Sγ	Cys39	$0.025\,(5.48 \times 10^{-2})$	$0.043\,(4.89 \times 10^{-2})$	$0.115\,(3.97 \times 10^{-2})$
Sγ	Cys49	$0.267\,(2.60 \times 10^{-1})$	$0.227\,(9.53 \times 10^{-2})$	$0.165\,(1.60 \times 10^{-2})$
C	Asn55	$0.070\,(2.05 \times 10^{-1})$	$0.029\,(8.48 \times 10^{-2})$	$0.015\,(4.24 \times 10^{-2})$
OXT	Asn55	$0.060\,(2.18 \times 10^{-1})$	$0.055\,(2.22 \times 10^{-1})$	$0.033\,(1.34 \times 10^{-1})$

[a]Fukui indices (in units of e) $q_A{}^-$ on the Zn^{2+}-coordinating thiolates and all other atoms with Fukui index larger than 0.050.
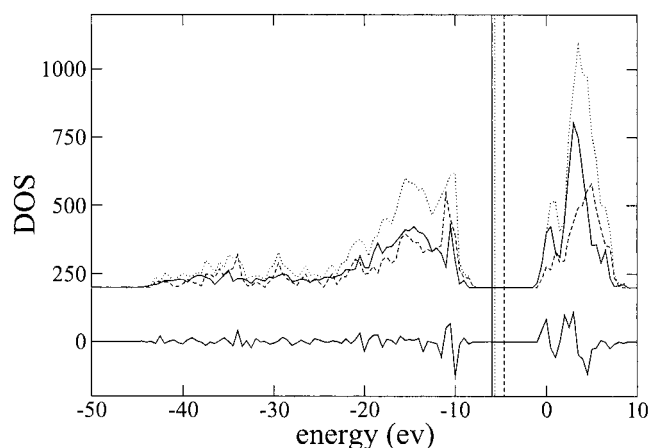

Fig. 5. **Top lines:** Density of states (DOS) of NC (solid lines), SL3-RNA (dashed lines), and the complex (dotted lines). **Bottom line:** The difference in DOS between the sum of the isolated species and the complexed one. The solid, dashed, and dotted vertical lines represent the Fermi levels for the protein, RNA, and complex, respectively. The plot is based on the optimized structure of the first entry of the NMR ensemble (PDB ID: 1A1T[52]). The bin size used for the plot is 0.5 eV.

using the SL3-RNA bound conformation.[31] It also correlates well with the experimental findings that the C-terminal zinc finger is more prone to electrophilic attack than the N-terminal zinc finger.[69–71] These results demonstrate that the approximate local hardness map may be used to reveal chemically reactive sites in biological macromolecules.

Density of Electronic States

In Figure 5, the upper plot shows the electronic density of states (DOS) for NC, SL3-RNA, and the protein-RNA complex, while the lower part shows the difference between the DOS of the complex and the sum of the DOS on the protein and RNA. It can be seen that the structure of the DOS differs for the separated protein, RNA, and their complex. The occupied orbitals of the RNA and the complex are shifted toward higher energy relative to those of the protein. Table IV compares the frontier orbital energies and orbital decompositions for these three molecules.

TABLE IV. Frontier Orbitals in NC, SL3-RNA and Their Complex

Orbital	HOMO			LUMO		
	Protein	RNA	Complex	Protein	RNA	Complex
Energy (eV)	−8.555	−7.751	−7.834	−0.614	−0.365	−0.789
Occupation[a]	1.565	1.941	1.934	1.409	1.813	1.996
Location[b]	Cys49	Gua13	Gua13	Arg32	Cyt19	Arg7

[a]Fractional orbital occupation $\Sigma_\mu |C_{i\mu}{}^\alpha|^2$, where the superscript α refers to the subsystem index, and subscript $i\mu$ refers to orbital i and atom μ.[51]
[b]This is the subsystem (amino acid or nucleotide residue) where the orbital is localized.

Due to the weak orbital interaction, the HOMO of the complex is expected to strongly resemble the HOMO of the isolated RNA, since it lies much higher than the one in the isolated protein. This is indeed the case. The HOMO in the complex remains localized on Gua13 and has slightly lower energy.

A further look at the energetics and orbital decomposition reveals that most of the high-lying orbitals are localized on the Zn^{2+}-coordinating Cys residues, including Cys15, Cys18, Cys28, Cys36, Cys39, and Cys49, with the predominant orbital coefficients on Sγ atoms (data not shown). These residues are expected to be most reactive toward small-molecule electrophiles. An exception is the second highest orbital ($\epsilon = -8.609\ eV$), which resides on Trp37, with the electronic density extending in the hetero-aromatic ring region (data not shown). The fact that the HOMO ($\epsilon = -8.555\ eV$) resides on Cys49 with the prevailing orbital coefficient on the thiolate, coincides with our previous prediction from the Fukui function analysis that Cys49 thiolate is the most reactive atom toward electrophilic attack.

The role of solvation has a large effect on the electronic density of states.[75,76] Here the focus has been on the effect of RNA complexation on the density of states of the isolated (uncomplexed) molecules. In the gas phase, the electronic states and Fermi levels of the uncomplexed molecules would be expected to shift dramatically due to the large difference in charge between NC and SL3-RNA. The effect of solvation is to greatly reduce this shift since

TABLE V. Relative Proton Potentials in OMTKY3

Residue	pK_a (expt)[a]	pK_a (calc)[b]	$\Delta\phi$ (V)[c]	$\Delta\phi_{sol}$ (V)[d]
Asp27	<2.3	3.4	−0.840	−3.096
Asp7	<2.7	3.3	−1.706	−3.128
Glu19	3.2	2.8	−1.521	−3.669
Glu10	4.2	3.5	−1.702	−3.070
Glu43	4.8	4.4	−3.277	−1.915

[a]Experimental pK_a values of the acidic groups in OMTKY3.[79]
[b]Calculated pK_a values[35] averaged over the NMR structure ensemble (PDB ID:1OMU[63]).
[c]Calculated relative proton electrostatic potentials.
[d]Calculated relative proton potentials due to the solvent reaction field.

the favorable electrostatic interaction of the complex is replaced by the solvent stabilization.

Relative Proton Potential

The pK_a values of ionizable groups in biomolecules are predominantly determined by their local electrostatic environment. The calculation of pK_a shifts in proteins requires consideration of an ensemble of protonation states in solution (including, in principle, conformational equilibration), and has been the subject of numerous theoretical investigations based on molecular mechanical force fields.[77,78] The electrostatic potential at a titratable proton site for a given protonation state can serve as an approximate measure of the relative proton affinity and hence is useful as a molecular descriptor. For weakly-coupled protonation states, the relative proton affinities are often correlated with corresponding pK_a values, and can be used to predict their order. A major advantage of using the proton potential as a pK_a descriptor lies in its low computational cost since it can be obtained from a linear-scaling electronic structure calculation without additional effort. Turkey ovomucoid third domain (OMTKY3) was chosen here as a benchmark protein to test the correlation between the experimental pK_a's[79] and the relative proton potentials. OMTKY3 contains multiple titratable groups and has been the subject of several theoretical studies.[35,80,81]

Table V compares the relative proton potentials at the carboxylic acid proton sites of OMTKY3 calculated using the linear-scaling quantum methods with theoretical[35] and experimental[79] pK_a values. The trend in the pK_α series is most closely followed by the relative proton potentials. Decreasing pK_a value (or increasing acidity) is correlated with less negative proton potential. The proton with the least negative potential is least stabilized and thus requires the smallest amount of energy to remove. The large pK_a value on Glu43 results primarily from decreased solvent stabilization of its ionized state (Table V). A plot of the relative proton potentials vs. experimental pK_a values reveals a nearly linear relation with a linear correlation coefficient $r_c = -0.996$, when Asp7 and Glu10 are excluded (Fig. 6). Asp7 and Glu10 are highly correlated: the CHARMM optimized structure shows that they are in close proximity and form hydrogen bonds upon protonation. Therefore, the electrostatic potentials at the acidic proton sites of Asp7 and Glu10 are highly coupled.
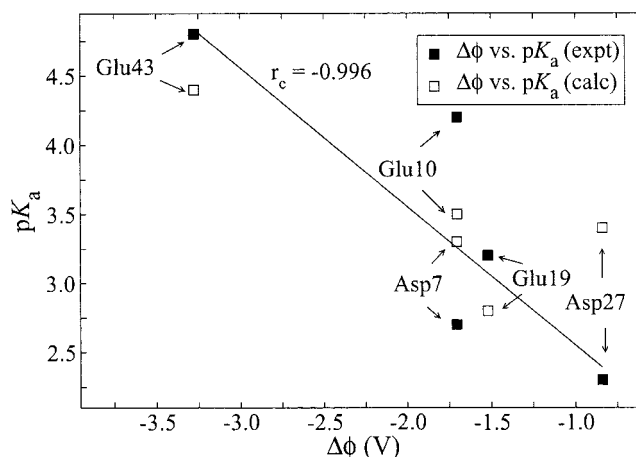

Fig. 6.   Correlation between the relative proton potentials $\Delta\Phi(V)$ at the carboxylic sites and the experimental and calculated pK_α values in OMTKY3. A linear regression line ($r_c = -0.996$) is drawn for $\Delta\Phi$ vs. pK_a (expt) excluding the highly correlated groups Asp7 and Glu10.

This dependency is not accounted for by the simplistic approach of computing relative proton potentials, which does not consider the ensemble of protonation states and conformational flexibility. Both the accurate experimental determination and theoretical prediction of pK_a's remain a challenging area. It is noteworthy that despite the large discrepancy between the estimated values from the linear regression and experimental pK_α's of Asp7 and Glu10, there is a fairly close agreement with other theoretical values (Fig. 6).

Nevertheless, the relative proton potential is a very useful quantity for predicting the order of pK_a's of weakly coupled ionizable groups and is particular suitable for identification of abnormally shifted pK_a's. A previous computational study[31] on NC was able to link the unusually high pK_a value on the Cys49 thiol[36] to its most negative proton potential. Figure 7 shows a plot of the calculated relative proton potentials (data taken from Khandogin et al.[31]) vs. the experimental pK_a values[36] of the protonated cysteine and histidine groups in the C-terminal zinc finger of HIV-1 NC. Again, a nearly linear relation is found between the relative proton potential and pK_a's with a correlation coefficient of −0.981. The correlation is somewhat less satisfactory at Cys36 and Cys39 due to their spatial proximity. These results clearly demonstrate that the relative electrostatic potentials at acidic proton sites can be employed as a molecular descriptor for qualitative prediction of the order of pK_a's in weakly coupled multiple titratable sites of a protein.

CONCLUSION

In this report, a set of macromolecular descriptors for the characterization of electrostatic and chemical features of solvated biological macromolecules has been described. The macromolecular descriptors can, in principle, be calculated using any electronic structure/solvation method. However, due to current constraints on computational resources, linear-scaling semi-empirical methods are
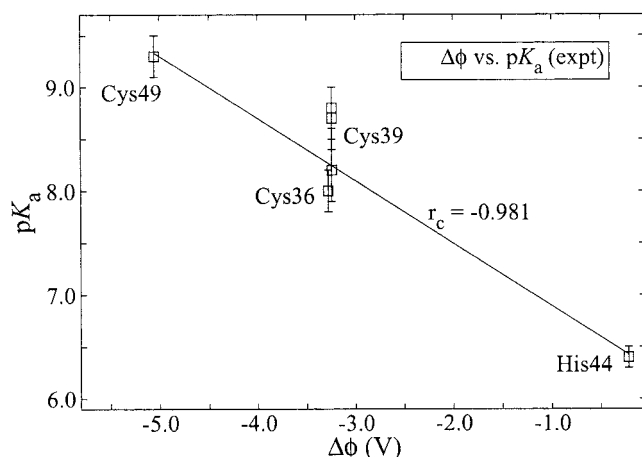
Fig. 7. Correlation between the relative proton potentials $\Delta\Phi(V)$ at the protonated C-terminal zinc-binding sites of NC and the experimental pKa's [36]. Experimental errors are indicated as error bars. For Cys36, all three measured pKa's are shown. A linear regression line ($r_c = -0.981$) is drawn for (expt) where for $\Delta\Phi$ vs. pKa (expt), where for Cys36 the average experimental pKa's is used.

among the most computationally feasible approaches. Development of improved semi-empirical models is undergoing a rebirth of interest due in part to their applications in hybrid quantum mechanical and molecular mechanical simulations and linear-scaling quantum calculations. The present work presents details of the derivation of the macromolecular descriptors and the implementation within the linear-scaling D&C semi-empirical program[29,51] and demonstrates their utility and predictive capabilities in applications to large biological systems in solution.

The electrostatic potential is calculated for a phosphate-binding protein mutant in both the free and ligand-bound forms, and the resulting ESP maps (AM1/COSMO or PM3/COSMO maps) are compared with those obtained from the PB calculations with classical atomic charges. The ESP maps derived from the AM1/COSMO method are in qualitative agreement with those from the PB method using the PARSE charges and radii. The main difference between the AM1/COSMO and the PARSE/PB maps is attributed to the form and magnitude of the solute charge representation that give rise to a "smoother" potential on the AM1/COSMO map. Further differences are illustrated by the example of the NC protein where the AM1/COSMO map uniquely distinguishes the electrostatic potentials of the two zinc finger regions.

The solvent-polarized atomic charges are calculated for both the NC protein and the protein-SL3 RNA complex. The Cys49 thiolate is found to be least stable among all Zn^{2+}-coordinating thiolates, as a result of its local electrostatic and solvent environment. RNA binding induced a net electron transfer by about 0.5 e from the RNA to protein. The Fukui indices and local hardness maps allow the prediction of chemical reactivity. The calculations on the NC protein reveal the Cys49 thiolate to be most chemically reactive, providing a rationale for its particularly high susceptibility toward electrophiles.[69-71] The approximate local hardness presented as a complementary

quantity to the Fukui index reveals the location of the HOMO density on a molecular surface. A detailed frontier orbital analysis confirms that the high-lying occupied orbitals are localized at the Zn^{2+}-coordinating cysteines with the predominant orbital coefficients on the thiolate atoms. In particular, the HOMO originates from the orbitals on Cys49. The orbital analysis confirms the reactivity predictions based on the Fukui indices and local hardness map.

Finally, the relative proton potential has been demonstrated to be a useful macromolecular descriptor for the estimation of the order in pKa's of weakly correlated titratable groups in biological macromolecules. A nearly linear relation between the relative proton potentials and the experimental pKa's is found for most acidic residues in OMTKY3 (except for the highly correlated groups). A plot of the relative proton potential at the protonated C-terminal zinc-binding sites of NC vs. the experimental pKa's also reveals a nearly linear relation. The macromolecular descriptors derivable from linear-scaling semi-empirical calculations provide detailed insight into the local electrostatic and chemical features of biological macromolecules; the relatively low computational cost makes the method particularly attractive for use in structure-based drug design.

REFERENCES

1. Gane PJ, Dean PM. Recent advances in structure-based rational drug design. Curr Opin Struct Biol 2000;10:401–404.
2. Klebe G. Recent developments in structure-based drug design. J Mol Med 2000;78:269–281.
3. Ramakrishnan V, Moore PB. Atomic structures at last: the ribosome in 2000. Curr Opin Struct Biol 2001;11:144–154.
4. Babine RE, Bender SL. Molecular recognition of protein-ligand complexes: applications to drug design. Chem Rev 1997;97:1359–1472.
5. Sotriffer C, Klebe G. Identi cation and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. Il Farmaco 2002;57:243–251.
6. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol 2002;6:439–448.
7. Chin K, Sharp KA, Honig B, Pyle AM. Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. Nature Struct Biol 1999;6:1055–1061.
8. Lee LP, Tidor B. Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. Protein Sci 2001;10:362–377.
9. Nicholls A, Honig B. A rapid nite difference algorithm, utilizing successive overrelaxation to solve the Poisson-Boltzmann equation. J Comput Chem 1991;12:435–445.
10. Luty BA, Davis ME, McCammon JA. Solving the nite-difference non-linear Poisson-Boltzmann equation. J Comput Chem 1992;13:1114–1118.
11. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A,

Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput Phys Commun 1995;91:57–95.

12. Im W, Beglov D, Roux B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Comput Phys Commun 1998;111:59–75.

13. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. Science 1995;268:1144–1149.

14. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. J Phys Chem 1994;98:1978–1988.

15. Rice WG, Schaeffer CA, Graham L, Bu M, McDougal JS, Orloff SL, Villinger F, Young M, Oroszlan S, Fesen MR, Pommier Y, Mendeleyev J, Kun E. The site of antiviral action of 3-nitrosobenzamide on the infectivity process of human immunode ciency virus in human lymphocytes. Proc Natl Acad Sci USA 1993;90:9721–9724.

16. Murray JS, Peralta-Inga Z, Politzer P, Ekanayake K, Lebreton P. Computational characterization of nucleotide bases: molecular surface electrostatic potentials and local ionization energies, and local polarization energies. Int J Quant Chem 2001;83:245–254.

17. Politzer P, Murray JS, Peralta-Inga Z. Molecular surface electrostatic potentials in relation to noncovalent interactions in biological systems. Int J Quant Chem 2001;85:676–684.

18. Head-Gordon M. Quantum chemistry and molecular processes. J Phys Chem 1996;100:13213–13225.

19. Yang W, P'érez-Jord'à J M. Linear scaling methods for electronic structure calculation. In: von Schleyer PR, editor. Encyclopedia of computational chemistry. New York: John Wiley and Sons; 1998. p 1496–1513.

20. Goedecker S. Linear scaling electronic structure methods. Rev Mod Phys 1999;71:1085–1123.

21. Scuseria GE. Linear scaling density functional calculations with Gaussian orbitals. J Phys Chem A 1999;103:4782–4790.

22. Gogonea V, Su'árez D, van der Vaart A, Merz KM Jr. New developments in applying quantum mechanics to proteins. Curr Opin Struct Biol 2001;11:217–223.

23. Galli G. Large-scale electronic structure calculations using linear scaling methods. Phys Status Solid B 2000;217:231–249.

24. Goedecker S, Scuseria GE. Linear scaling electronic structure methods in chemistry and physics. IEEE Comput Sci Eng 2003;5: 14–21.

25. Nadig G, Van Zant LC, Dixon SL, Merz KM Jr. Charge-transfer interactions in macromolecular systems: a new view of the protein/water interface. J Am Chem Soc 1998;120:5593–5594.

26. Lewis JP, Liu S, Lee T-S, Yang W. A linear-scaling quantum mechanical investigation of cytidine deaminase. J Comput Phys 1999;151:242–263.

27. Khandogin J, Hu A, York DM. Electronic structure properties of solvated biomolecules: a quantum approach for macromolecular characterization. J Comput Chem 2000;21:1562–1571.

28. Klamt A, Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 1993;2 2:799–805.

29. York DM, Lee T-S, Yang W. Parameterization and ef cient implementation of a solvent model for linear-scaling semiempirical quantum mechanical calculations of biological macromolecules. Chem Phys Lett 1996;263:297–306.

30. York DM, Karplus M. A smooth solvation potential based on the conductor-like screening model. J Phys Chem A 1999;103:11060–11079.

31. Khandogin J, Musier-Forsyth K, York DM. Insights into the regioselectivity and RNA-binding af nity of HIV-1 nucleocapsid protein from linear-scaling quantum methods. J Mol Biol 2003;330: 993–1004.

32. Matthew JB, Gurd FR, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. CRC Crit Rev Biochem 1985;18:91–197.

33. Murray JS, Brinck T, Grice ME, Politzer P. Correlations between molecular electrostatic potentials and some experimentally-based indices of reactivity. J Mol Struct (Theochem) 1992;256:29–45.

34. Dimitrova V, Ilieva S, Galabov B. Electrostatic potential at atomic sites as a reactivity descriptor for hydrogen bonding. complexes of monosubstituted acetylenes and ammonia. J Phys Chem A 2002; 106:11801–11805.

35. Forsyth WR, Gilson MK, Antosiewicz J, Jaren OR, Robertson AD. Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. Biochemistry 1998;37:8643–8652.

36. Bombarda E, Morelleta N, Cherradi H, Spiess B, Bouaziz S, Grell E, Roques BP, M'ély Y. Determination of the pK of the four Zn-coordinating residues of the distal nger motif of the HIV-1 nucleocapsid protein: consequences on the binding of Zn . J Mol Biol 2001;310:659–672.

37. Vuilleumier R, Sprik M. Electronic properties of hard and soft ions in solution: Aqueous Na and Ag compared. J Chem Phys 2001;115: 3454–3468.

38. Langenaeker W, De Proft F, Geerlings P. Development of local hardness Related reactivity indices: their application in a study of the S at Monosubstituted Benzenes within the HSAB context. J Phys Chem 1995;99:6424–6431.

39. Roy RK, Krishnamurti S, Geerlings P, Pal S. Local softness and hardness based reactivity descriptors for predicting intra- and intermolecular reactivity sequences: carbonyl compounds. J Phys Chem A 1998;102:3746–3755.

40. Geerlings P, de Proft F. HSAB principle: Applications of its global and local forms in organic chemistry. Int J Quant Chem 2000;80: 227–235.

41. Zuloaga F, Domard M, Pautet F, Fillion H, Tapia R. Diels Alder reactions of a thiazole o-quinodimethane with 2- and 3-bromo-5-hydroxynaphthoquinones: a theoretical study. Tetrahedron 2000; 56:1701–1706.

42. Sivanesan D, Subramanian V, Unni Nair B. Quanti cation of reactive sites in DNA bases using condensed Fukui functions. J Mol Struct (Theochem) 2001;544:123–139.

43. Maynard AT, Huang M, Rice WG, Covell DG. Reactivity of the HIV-1 nucleocapsid protein p7 zinc nger domains from the perspective of density-functional theory. Proc Natl Acad Sci USA 1998;95: 11578–11583.

44. Khandogin J, Gregersen BA, York DM. A smooth solvation method for d-orbital semiempirical calculations of biological reactions: application to RNA transphosphorylation. J Phys Chem B submitted, 2004.

45. Li J, Zhu T, Cramer CJ, Truhlar DG. New class IV charge model for extracting accurate partial charges from wave functions. J Phys Chem A 1998;102:1820–1831.

46. Kaminski GA, Jorgensen WL. A quantum mechanical and molecular mechanical method based on CM1A charges: applications to solvent effects on organic equilibria and reactions. J Phys Chem B 1998;102:1787–1796.

47. Li J, Cramer CJ, Truhlar DG. Application of a universal solvation model to nucleic acid bases: Comparison of semiempirical molecular orbital theory, ab initio Hartree-Fock theory, and density functional theory. Biophys Chem 1999;78:147–155.

48. Gogonea V, Merz K Jr. Fully quantum mechanical description of proteins in solution. Combining linear scaling quantum mechanical methodologies with the Poisson-Boltzmann equation. J Phys Chem A 1999;103:5171–5188.

49. Khandogin J, York DM. Quantum mechanical characterization of nucleic acids in solution: a linear-scaling study of charge uctuations in DNA and RNA. J Phys Chem B 2002;106:7693–7703.

50. Pearson RG. Chemical hardness. Weinheim: Wiley-VCH; 1997.

51. Lee T-S, York DM, Yang W. Linear-scaling semiempirical quantum calculations for macromolecules. J Chem Phys 1996;105:2744–2750.

52. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 nucleocapsid protein bound to the SL3-RNA recognition element. Science 1998;279:384–388.

53. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 1983;4: 187–217.

54. MacKerell AD Jr, Brooks B, Brooks CL III, Nilsson L, Roux B, Won Y, Karplus M. CHARMM: the energy function and its parameterization with an overview of the program. In: v. R. Schleyer P, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR, editors. Encyclopedia of computational chemistry, vol. 1. Chichester, UK: John Wiley & Sons; 1998. p 271–277.

55. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B,

Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wi'órkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

56. Foloppe N, MacKerell AD Jr. All-atom empirical force eld for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 2000;21:86–104.

57. Stote RH, Karplus M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. Proteins 1995;23: 12–31.

58. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.

59. Amarasinghe GK, De Guzman RN, Turner RB, Summers MF. NMR structure of stem-loop SL2 of the HIV-1 RNA packaging signal reveals a novel A-U-A base-triple platform. J Mol Biol 2000;299:145–156.

60. Ledvina PS, Yao N, Choudhary A, Quiocho FA. Negative electrostatic surface potential of protein sites speci c for anionic ligands. Proc Natl Acad Sci USA 1996;93:6786–6791.

61. Wang Z, Luecke H, Yao N, Quiocho FA. A low energy short hydrogen bond in very high resolution structures of protein receptor phosphate complexes. Nature Struct Biol 1997;4:519–523.

62. Thiel W. Perspectives on semiempirical molecular orbital theory. In: Prigogine I, Rice SA, editors. Adv Chem Phys, vol. 93. New York: John Wiley and Sons; 1996. p 703–757.

63. Hoogstraten CG, Choe S, Westler WM, Markley JL. Comparison of the accuracy of protein solution structures derived from conventional and network-edited NOESY data. Protein Sci 1995;4:2289–2299.

64. York D, Lee T-S, Yang W. Quantum mechanical study of aqueous polarization effects on biological macromolecules. J Am Chem Soc 1996;118:10940–10941.

65. Wang BZ, Ford GP. New approach to the rapid semiempirical calculation of molecular electrostatic potentials based on the AM1 wave function-comparison with ab initio HF/6-31G* results. J Comput Chem 1993;14:1101–1111.

66. Bakowies D, Thiel W. Semiempirical treatment of electrostatic potentials and partial charges in combined quantum mechanical and molecular mechanical approaches. J Comput Chem 1996;17: 87–108.

67. Clark T. Quo vadis semiempirical MO-theory? J Mol Struct (Theochem) 2000;530:1–10.

68. Maynard AT, Covell DG. Reactivity of zinc nger cores: analysis of protein packing and electrostatic screening. J Am Chem Soc 2001;123:1047–1058.

69. Hathout Y, Fabris D, Han MS, Sowder R II, Henderson LE, Fenselau C. Characterization of intermediates in the oxidation of zinc ngers in human immunodeficiency virus type 1 nucleocapsid protein P7. Drug Metab Dispos 1997;24:1395–1400.

70. Chertova EN, Kane BP, McGrath C, Johnson DG, Sowder RC II, Arthur LO, Henderson LE. Probing the topography of HIV-1 nucleocapsid protein with the alkylating agent N-Ethylmaleimide. Biochemistry 1998;37:17890–17897.

71. Basrur V, Song Y, Mazur SJ, Higashimoto Y, Turpin JA, Rice WG, Inmani JK, Appella E. Inactivation of HIV-1 nucleocapsid protein p7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action. J Biol Chem 2000;275: 14890–14897.

72. Michalak A, Proft FD, Geerlings P, Nalewajski RF. Fukui functions from the relaxed kohn-sham orbitals. J Phys Chem A 1999;103:762–771.

73. Arulmozhiraja S, Kolandaivel P. Condensed Fukui function: dependency on atomic charges. Mol Phys 1997;90:55–62.

74. Parr RG, Pearson RG. Absolute hardness: companion parameter to absolute electronegativity. J Am Chem Soc 1983;105:7512–7516.

75. York DM, Lee T-S, Yang W. Quantum mechanical treatment of biological macromolecules in solution using linear-scaling electronic structure methods. Phys Rev Lett 1998;80:5011–5014.

76. York DM. Application of linear-scaling electronic structure methods to the study of polarization of proteins and DNA in solution. In: Gao J, Thompson M, editors. Combined quantum mechanical and molecular mechanical methods. ACS Symposium Series 712. New York: Oxford University Press; 1998. p 275–287.

77. Bashford D, Karplus M. pK s of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.

78. Yang A-S, Gunner MR, Sampogna R, Sharp K, Honig B. On the calculation of pK's in proteins. Proteins 1993;15:252–265.

79. Schaller W, Robertson AD. pH, ionic strength, and temperature dependences of ion-ization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 1995;34:4714–4723.

80. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK 's in proteins. Biochemistry 1996;35:7819–7833.

81. Li H, Hains AW, Everts JE, Robertson AD, Jensen JH. The prediction of protein p 's using QM/MM: the p of lysine 55 in turkey ovomucoid third domain. J Phys Chem B 2002;106:3486–3494.

82. Senet P. Chemical hardnesses of atoms and molecules from frontier orbitals. Chem Phys Lett 1997;275:527–532.

83. Yang W, Parr RG. Hardness, softness, and the fukui function in the electron theory of metals and catalysis. Proc Natl Acad Sci 1985;82:6723–6726.

84. Parr RG, Yang W. Density functional approach to the frontier-electron theory of chemical reactivity. J Am Chem Soc 1984;106: 4049–4050.

85. Szabo A, Ostlund NS. Modern quantum chemistry: introduction to advanced electronic structure theory, 1st ed.. New York: Dover Publications, Inc. 1996.

86. Mulliken RS. Electronic population analysis on LCAO-MO: linear combination of atomic orbital-molecular orbital molecular wave functions. I. J Chem Phys 1955;23:1833–1840.

87. Mulliken RS. Electronic population analysis on LCAO-MO: linear combination of atomic orbital-molecular orbital molecular wave functions. II. Overlap populations, bond orders, and covalent bond energies. J Chem Phys 1955;23:1841–1846.

88. Ghosh SK, Berkowitz M. A classical uid-like approach to the density functional formalism of many-electron systems. J Chem Phys 1985;83:2976–2983.

89. Berkowitz M, Parr RG. Molecular hardness and softness, local hardness and softness, hardness and softness kernels, and relations among these quantities. J Chem Phys 1988;88:2554–2557.

90. Chermette H. Chemical reactivity indexes in density functional theory. J Comput Chem 1999;20:129–154.

APPENDIX

In density-functional theory, the electronic energy is written as a functional of the electron density $\rho(\boldsymbol{r})$ in the form

$$E[\rho] = F[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d^3r \qquad (1)$$

where $v(\boldsymbol{r})$ is the external potential (for example, the nuclear potential in the case of molecules in the absence of applied fields), and $F[\rho]$ is a universal functional of the density only, and consists of the electronic kinetic energy $T[\rho]$ and the electron-electron interation energy $V_{ee}[\rho]$. With reference to the commonly applied Kohn-Sham formalism, $F[\rho]$ is written as the sum of the non-interacting kinetic energy $T_s[\rho]$, classical electrostatic energy $J[\rho]$, and the exchange-correlation energy $K[\rho]$:$F[\rho] = T[\rho] + V_{ee}[\rho] = T_s[\rho] + J[\rho] + K[\rho]$. Of particular interest in the present work is the classical electrostatic energy, which is given by

$$J[\rho] = \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d^3r d^3r' \qquad (2)$$

The ground-state electron density $\rho(\boldsymbol{r})$ and external potential $v(\boldsymbol{r})$ are related by the constrained variational condition

$$\delta\{E[\rho] - \mu(\int \rho(\mathbf{r})d^3r - N)\} = 0 \qquad (3)$$

where μ is a Lagrange multiplier on the normalization constraint $\int \rho(\mathbf{r})d^3r = N$. Eq. 3 establishes a mapping (to within a trivial constant) between the ground-state electron density and external potential for a given number of electrons: $\rho \leftrightarrow$. From Eqs. 1 and 3 it can be seen that the chemical potential can be written as

$$\mu = \left[\frac{\delta E}{\delta\rho(\mathbf{r})}\right]_v = \left[\frac{\partial E}{\partial N}\right]_v \qquad (4)$$

where it is understood that the functional derivative is evaluated at the ground-state density (a convention that is adopted throughout the report). The chemical potential is hence the first derivative of the energy with respect to the number of electrons, and is related to minus the Mulliken electronegativity $\chi = 1/2(IP + EA)$, where IP is the ionization potential and EA is the electron affinity, by a finite difference approximation. The corresponding second derivative (curvature) of the energy with respect to the number of electrons is called the chemical hardness or the global hardness[74]

$$\eta = \left[\frac{\partial\mu}{\partial N}\right]_v = \left[\frac{\partial^2 E}{\partial N^2}\right]_v \qquad (5)$$

The above equation does not include the factor of 1/2 in the definition of η as used by some authors. The chemical hardness is routinely approximated by finite differences, in analogy to the chemical potential, as $\eta = IP - EA$.[50,82] The inverse of chemical hardness is called the chemical softness or the global softness

$$S = \frac{1}{\eta} = \left[\frac{\partial N}{\partial\mu}\right]_v \qquad (6)$$

Two related quantities for measuring regioselectivity are the *local* softness and hardness. The local softness is defined as[83]

$$s(\mathbf{r}) = \left[\frac{\partial\rho(\mathbf{r})}{\partial\mu}\right]_v = \left[\frac{\partial\rho(\mathbf{r})}{\partial N}\right]_v \left[\frac{\partial N}{\partial\mu}\right]_v = \left[\frac{\partial\rho(\mathbf{r})}{\partial N}\right]_v \cdot S \qquad (7)$$

and integrates to the global softness S

$$\int s(\mathbf{r})d^3r = S \qquad (8)$$

In Eq. 7 the term in front of S is referred to as the Fukui function $f(\mathbf{r})$[84]

$$f(\mathbf{r}) = \left[\frac{\partial\rho(\mathbf{r})}{\partial N}\right]_v = \left[\frac{\delta\mu}{\delta\upsilon(\mathbf{r})}\right]_N = \frac{s(\mathbf{r})}{S} \qquad (9)$$

The Fukui function contains the same information as the local softness (with the exception of the normalization), and will be the focus of later discussions. In the Kohn-Sham (KS) density-functional theory, the electron density can be expressed in terms of the KS molecular orbitals as

$$\rho(\mathbf{r}) = \sum_i^N n_i\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}) \qquad (10)$$

where ψ_i is the KS spin orbital and η_i is the orbital occupation number: 1 for occupied and 0 for virtual orbitals. Inserting the above expression into Eq. 10 results in the following equation for the calculation of Fukui function

$$f(\mathbf{r}) = \sum_i^N \frac{\partial n_i}{\partial N}|\psi_i|^2 + \sum_i^N n_i \frac{\partial\psi_i}{\partial N}\psi_i^* + \sum_i^N n_i\psi_i \frac{\partial\psi_i^*}{\partial N} \qquad (11)$$

where the first term represents the contribution due to the change in the occupation number only and corresponds to the Fukui function under the frozen orbital approximation, and the last two terms represent the contribution due to the orbital relaxation. To simplify the calculations of the large biological macromolecules considered here, the orbital relaxation component was neglected in accord with Koopmans' theorem.[85]

To circumvent the formal discontinuity in the derivative of the Fukui function for integer number of electrons in the zero temperature limit, and to produce a more robust and more delocalized reactivity index, the Fukui function is approximated by finite difference as

$$f(\mathbf{r})^\pm \simeq \pm\frac{1}{\Delta N}[\rho(\mathbf{r};N \pm \Delta N) - \rho(\mathbf{r};N)]$$

$$f(\mathbf{r})^0 \simeq \frac{1}{2\Delta N}[\rho(\mathbf{r};N + \Delta N) - \rho(\mathbf{r};N - \Delta N)] = \frac{1}{2}[f(\mathbf{r})^+ + f(\mathbf{r})^-]$$

$$(12)$$

where $f(\mathbf{r})^-$, $f(\mathbf{r})^+$, and $f(\mathbf{r})^0$ are the left, right, and mean Fukui functions. In the case that the finite difference step size ΔN is 1, Fukui function under the frozen orbital approximation can be written as

$$f(\mathbf{r})^- \simeq |\psi_{HOMO}|^2$$

$$f(\mathbf{r})^+ \simeq |\psi_{LUMO}|^2$$

$$f(\mathbf{r})^0 \simeq \frac{1}{2}[|\psi_{HOMO}|^2 + \psi_{LUMO}|^2] \qquad (13)$$

where ψ_{HOMO} are the highest occupied and ψ_{LUMO} are the lowest occupied KS orbitals. The above equations suggest the left, right, and mean Fukui functions to be interpreted as the HOMO, LUMO, and the average of both densities, respectively. Thus, $f(\mathbf{r})^-$, $f(\mathbf{r})^+$, and $f(\mathbf{r})^0$ can serve as chemical reactivity indices for electrophilic, nucleophilic, and radical attacks, respectively. With a choice of $\Delta N = 1$, however, the large number of closely degenerate eigenstates of a macromolecular system makes approximate Fukui index very sensitive to subtle environmental changes. To obviate this difficulty, several values of ΔN that scale in proportion with the system size were tested and found to be more robust and provide a better physical picture of the regioselective reactivity. In order to assign spatial locality, it is often times convenient to use the

condensed Fukui functions, q_k^+, q_k^-, q_k^0, obtained by applying an electronic density partitioning scheme, such as those used to derive Mulliken[86,87] or CM2[45] charges.

A globally reciprocal quantity to the local softness is the local hardness:

$$\int s(\mathbf{r})\eta(\mathbf{r})d^3r = 1 \qquad (14)$$

By analogy to Eq. 7, one possible definition of the local hardness is given as[88]

$$\eta(\mathbf{r}) = \left[\frac{\delta\mu}{\delta\rho(\mathbf{r})}\right]_v \qquad (15)$$

However, according to the Hohenberg-Kohn theorem, which was later generalized by Levy, v is completely determined by ρ for a stationary state, and hence the definition (Eq. 15) is ambiguous. A general form of local hardness can be written as[88]

$$\eta_\lambda(\mathbf{r}) = \int \eta(\mathbf{r},\mathbf{r}')\lambda(\rho(\mathbf{r}'))d^3r' \qquad (16)$$

where $\eta(\mathbf{r},\mathbf{r}')$ is the hardness kernel[89]

$$\eta(\mathbf{r},\mathbf{r}') = \frac{\delta^2 F[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \qquad (17)$$

$\lambda(\rho(\mathbf{r}))$ is a normalized composite function of ρ,

$$\int \lambda(\rho(\mathbf{r}))d^3r = 1 \qquad (18)$$

and $F[\rho]$ is the universal functional of Eq. 3. A common simplification of the hardness kernel is to neglect the contributions from the nonclassical kinetic and exchange-correlation energies,[90] leading to

$$\eta(\mathbf{r},\mathbf{r}') \approx \frac{\delta^2 J[\rho]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} = \frac{1}{|\mathbf{r} - \mathbf{r}'|} \qquad (19)$$

where $J[\rho]$ is the classical electrostatic energy of Eq. 2. This approximation is reasonable for outer region of the molecule due to the exponential decay of the electron density, and has the advantage that it is independent of $\rho(\mathbf{r})$ and can be easily calculated.

A common convention is to use ρ/N as the composite function, resulting in the approximate local hardness $\eta_D(\mathbf{r})$,[88] where the subscript D stands for "density." $\eta_D(\mathbf{r})$ represents the electronic part of the molecular electrostatic potential; however, for the purposes of this work, there is little information gained from this quantity as it contains similar information as the solvent-polarized molecular electrostatic potential. Here, the Fukui function $f(\mathbf{r})$ is chosen as the composite function.[88–90] Combining Eq. 16 and Eq. 17, an approximate local hardness $\eta_F(\mathbf{r})$ (F stand for "Fukui") can be defined in terms of the Fukui function as

$$\eta_F(\mathbf{r}) \simeq \int \frac{f(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r' = \int \frac{[\partial\rho(\mathbf{r}')/\partial N]_v}{|\mathbf{r} - \mathbf{r}'|} d^3r'$$

$$= \frac{\partial}{\partial N}\left[\int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r'\right]_v = \left[\frac{\partial V_{el}(\mathbf{r})}{\partial N}\right]_v \qquad (20)$$

where $V_{el}(\mathbf{r})$ represents the electronic part of the molecular electrostatic potential. Thus, $\eta_F(\mathbf{r})$ can be considered as the frontier orbital electrostatic potential or as the change in the electronic part of the molecular electrostatic potential. In analogy to the left, right, and mean Fukui functions, there exist the left, right, and mean local hardness functions $\eta_F(\mathbf{r})^-$, $\eta_F(\mathbf{r})^+$, and $\eta_F(\mathbf{r})^0$ that represent the electrostatic potential due to the HOMO, LUMO and the average of both densities, respectively.

Another useful response property is the derivative of the total energy with respect to the nuclear charge Z_α, which, by making use of the Hellmann-Feynman theorem, can be written as:

$$\frac{\partial}{\partial Z_\alpha}(E[\rho] + V_{NN}) = <\Psi|\frac{\partial H}{\partial Z_\alpha}|\Psi> + \frac{\partial}{\partial Z_\alpha}\sum_{\alpha<\beta}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}$$

$$= -\int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_\alpha|} d^3r + \sum_{\beta\neq\alpha}\frac{Z_\beta}{|\mathbf{r}_\alpha - \mathbf{r}_\beta|} = \Phi_{tot}(\mathbf{r}_\alpha) \quad (21)$$

The quantity $\Phi_{tot}(\mathbf{R}_\alpha)$ represents the electrostatic potential at nucleus α due to all electrons and other nuclei. With solvation effects taken into account, $\Phi_{tot}(\mathbf{R}_\alpha)$ values at titratable proton positions are useful for assessing the relative proton affinities of these sites in biomolecules.