JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2018, 58, 1836−1846

pubs.acs.org/jcim

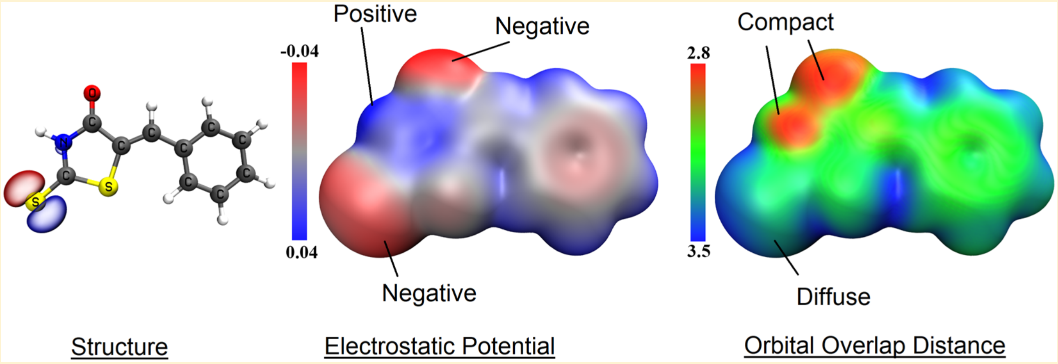
An Orbital-Overlap Complement to Ligand and Binding Site Electrostatic Potential Maps

Arshad Mehmood,[†] Stephanie I. Jones,[†] Peng Tao,[‡] and Benjamin G. Janesko*,[†]

[†]Department of Chemistry and Biochemistry, Texas Christian University, 2800 South University Drive, Fort Worth, Texas 76129, United States

[‡]Department of Chemistry, Southern Methodist University, P.O. Box 750314, Dallas, Texas 75275, United States

Supporting Information



ABSTRACT: Orbitals and orbital overlap are important concepts in chemistry but are seldom incorporated into medicinal chemistry analyses of drug−target interactions. Our orbital overlap distance $D(\mathbf{r})$ quantifies the size of the "test orbital" that best overlaps with a system's computed orbitals at point **r**. The overlap distance provides information about all of the occupied orbitals across a molecule, extending frontier orbital (Fukui) analysis and complementing widely used maps of the surface electrostatic potential. We present the first tests of the overlap distance for problems in medicinal chemistry. The overlap distance quantifies the different coordination chemistries of pairs of metal cations possessing similar charges and ionic radii. Combining the overlap distance and electrostatic potential provides a rich picture of the binding sites for chemically "hard" versus "soft" cations in formylglycine-generating enzyme and extends frontier orbital analysis in quantifying the chemistry of promiscuous binders. We conclude by showing how the electrostatic potential and overlap distance combine to give a novel and experimentally testable prediction for improving the in vivo activity of centromere-associated protein E inhibitors. The results motivate including the overlap distance alongside electrostatic potential maps in medicinal chemistry.

■ INTRODUCTION

Overview. Orbitals and orbital overlap are at the heart of chemistry. Orbital overlap in chemical bonds,[1] frontier orbital interactions,[2] isolobality,[3] and charge versus orbital control of reactivity[4] are central to chemical theory. Orbital overlap explains why phenyl sulfide is a better nucleophile than the more negatively charged phenyl oxide, why deprotonated amides perform nucleophilic attack at nitrogen rather than the more negatively charged oxygen,[5] and why cations of comparable charge and ionic radius can have strikingly different biochemistry.[6] Molecular orbital calculations are increasingly important in biological and medicinal chemistry, as illustrated in recent applications to quantitative structure−activity relationships (QSARs),[7] ligand−target interactions,[8] mechanistic proposals for enzyme catalysis,[9] and simulations of entire proteins.[10]

However, orbitals *as a conceptual tool* are used less in biological and medicinal chemistry. Classical analyses of one orbital at a time (e.g., highest occupied and lowest unoccupied orbitals[2]), though useful in some studies,[7,11] can be inefficient for large ligands and active sites. Instead, medicinal chemists often focus on electrostatic and hydrophobic interactions visualized across a biomolecule's entire surface.[12−15] This imbalance is illustrated in a recent study of 17a-hydroxylase-17,20-lyase inhibition, where sophisticated density functional theory (DFT) molecular orbital calculations were analyzed solely in terms of computed molecular electrostatic potentials (MESPs).[16] Many investigators have explored orbital-based descriptors in biochemistry.[17] Recent examples include condensed Fukui functions,[18] chemical hardness and softness,[19,20] and the electron localization function (ELF).[11] However, these powerful tools often have a steep learning curve and are arguably unfamiliar to many experimentalists.[21]

The Orbital Overlap Distance. Our orbital overlap distance is intended to more effectively bridge the divide

Received: June 11, 2018
Published: August 30, 2018

DOI: 10.1021/acs.jcim.8b00370
J. Chem. Inf. Model. 2018, 58, 1836−1846

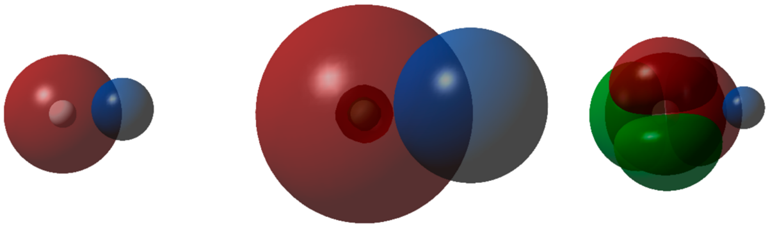


Figure 1. Calculation of the orbital overlap distance at a point **r** on the surface of helium (left), lithium (middle), and oxygen (right) atoms. The α-spin valence orbitals (He 1s, Li 2s, O 2s and 2p) are shown in red and green. The EDR test function $C_d \exp(-|\mathbf{r} - \mathbf{r}'|^2/d^2)$ centered at point **r** is shown in blue. The test function's *width d* is chosen to maximize its overlap with the occupied orbitals. The "orbital overlap distance" $D(\mathbf{r})$ is defined as that width.

between classical analyses of individual orbitals and visualization of chemically intuitive quantities on the entire surface of large biomolecules. The orbital overlap distance quantifies the size of the orbital lobes on an approaching ligand that best overlap with the occupied MOs of the molecule of interest. The orbital overlap distance is based on our electron delocalization range function, EDR($\mathbf{r}$; d).[22] The EDR quantifies the extent to which the molecular orbitals around point **r** overlap with a hydrogenic "test function" orbital of width d centered at **r**:

$$\mathrm{EDR}(\mathbf{r};\, d) = \rho^{-1/2}(\mathbf{r}) \int \mathrm{d}^3\mathbf{r}'\gamma(\mathbf{r},\, \mathbf{r}')\, \mathrm{C}_d \exp(-|\mathbf{r} - \mathbf{r}'|^2/d^2) \tag{1}$$

where $\gamma(\mathbf{r}, \mathbf{r}') = \sum_{i\sigma} n_{i\sigma}\psi_{i\sigma}(\mathbf{r})\psi_{i\sigma}(\mathbf{r}')$ is the one-particle density matrix constructed from all of the spin−orbitals $\psi_{i\sigma}(\mathbf{r})$ with nonzero occupancies $n_{i\sigma}$ and $C_d = [2/(\pi d^2)]^{3/4}$ is a factor that ensures that the value of the EDR is between −1 and +1. The orbital overlap distance $D(\mathbf{r})$ is defined as the value of d that maximizes EDR($\mathbf{r}$; d) at point **r**. Like the electrostatic potential, the overlap distance and EDR are functions of the full one-particle density matrix.[22] This makes them independent of orbital-localizing unitary transforms and readily evaluated from multideterminant and multireference ab initio calculations.

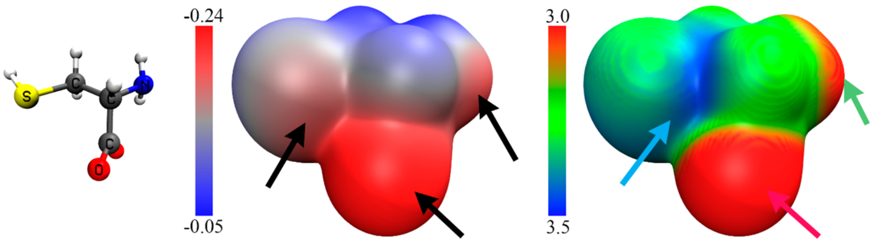
Figure 1 shows how we compute the orbital overlap distance $D(\mathbf{r})$ at a single point **r** on the surface of hydrogen, helium, and oxygen atoms. The green and red surfaces in Figure 1 show the occupied α-spin valence molecular orbitals. The blue surfaces are the EDR test function $C_d \exp(-|\mathbf{r} - \mathbf{r}'|^2/d^2)$ in eq 1, which is centered at point **r**. The orbital overlap distance $D(\mathbf{r})$ is the value of the test function width d that *maximizes the test function's overlap* with the molecular orbitals. The overlap between the relatively diffuse lithium atom 2s orbital and the EDR test function is maximized at a relatively large width, $D(\mathbf{r}) = 6.5$ bohr. The overlap between the more compact doubly occupied helium atom 1s orbital and the EDR test function is maximized at a smaller width, $D(\mathbf{r}) = 3.1$ bohr. The overlap between the oxygen atom 2s and 2p orbitals and the EDR test function is maximized when the test function overlaps with *one of the lobes* of the oxygen p orbital. This leads to a small width, $D(\mathbf{r}) = 2.6$ bohr.

Before continuing, we clarify two potentially confusing aspects of Figure 1. First, the density matrix formalism in eq 1 ensures that the orbital overlap distance is evaluated from ***all*** orbitals with nonzero occupancy. Such density-matrix-dependent quantities are independent of "localizing" or "hybridizing" transforms of integer-occupied orbitals. For example, the overlap distance and electrostatic potential computed for ethylene are independent of whether the canonical delocalized occupied molecular orbitals are transformed to "σ and π bonds" or "banana bonds". Second, the results for oxygen show that the orbital overlap distance characterizes the size of *individual orbital lobes*. This naturally corresponds to the valence-bond picture of bonding, in which, for example, ethane formation from two sp^3-hybridized H$_3$C* fragments involves overlap of the main lobes of the two singly occupied carbon sp^3 orbitals. Rather than choosing different test functions for different orbitals, we choose a single test function to overlap with a single orbital lobe.

The orbital overlap distance is freely available in the Multiwfn package[23] and is available in the Gaussian 16 *cubegen* utility.[24] Both of these packages compute the electron density, electrostatic potential, and overlap distance on a three-dimensional grid of points. These three-dimensional grids are used by visualization software (here the VMD package) to depict the electrostatic potential or overlap distance on particular density isosurfaces. These three-dimensional grids may also be used to evaluate atom-averaged values of the density (to determine atomic partial charges), electrostatic potential, or overlap distance.

The orbital overlap distance contains information that is qualitatively different from the information contained in the electrostatic potential. Molecular electrostatic potentials are evaluated from the nuclear positions and the electron probability density distribution $\rho(\mathbf{r}) = \gamma(\mathbf{r},\mathbf{r})$, i.e., from the diagonal part of the one-particle density matrix. The off-diagonal part of the density matrix, the part sampled by the EDR, is key to orbital overlap, electronic kinetic energy,[25] and bonding versus antibonding interactions.[26,27]

We have shown in previous work how the orbital overlap distance and EDR give quantitative chemically relevant predictions that are unavailable from electrostatics alone. System averages of the orbital overlap distance computed in 76 structurally diverse anionic water clusters show a quantitative one-to-one correlation with a very different measure of the "size" of the solvated electron, the radius of gyration of the singly occupied molecular orbital.[28,29] Such system averages also give quantitative agreement with Mollwo−Ivey relations for the "size" of electrons trapped in alkali halide bulk F-center defects[30] and for isolated atoms' highest occupied molecular orbital (HOMO) radii[31] and quantify the relationship between cavity size and trapped electron delocalization in high-pressure electrides.[32,33] Topological analysis of the EDR following Bader's approach[34] gives attractors whose locations in space are consistent with the topology of the ELF.[25,35,36] The orbital



Site	Value of surface electrostatic potential (au)	Value of surface orbital overlap distance D(r) (bohr)
O	-0.25	3.00
N	-0.18	3.15
S	-0.15	3.45

Figure 2. (top, left to right) Computed structure, molecular electrostatic potential, and overlap distance $D(\mathbf{r})$ of deprotonated cysteine. The electrostatic potential distinguishes the three Lewis basic sites (black arrows) from the surrounding amino acid. The overlap distance distinguishes the diffuse and chemically soft sulfur base (blue arrow) from the compact and chemically hard oxygen base (red arrow) and nitrogen base (green arrow). (bottom) Values of the electrostatic potential and overlap distance at the three points shown with arrows.

overlap distances computed at these attractors provide chemically reasonable "sizes" of core, bonding, and lone-pair orbitals; distinguish Au^+ cores from delocalized metallic electrons; and quantify the additional delocalization of stretched bonds for ammonia dissociation on Si(001).[35] The overlap distance and EDR computed from DFT calculations generally match high-level ab initio calculations.[31] *Differences* between EDRs computed from single- versus multireference calculations quantify the localizing effects of strong correlation in stretched H_2 molecule,[22] stretched carbon monoxide,[31] other stretched polar covalent bonds,[37] and a cluster model of lithium−ammonia solutions' concentration-dependent transition to a metallic state.[29,38] Atom-averaged orbital overlap distances rationalize several situations where atomic partial charges alone provide an incomplete picture of reactivity: why PhS^- is a better nucleophile than PhO^- in S_N2 reactions with MeI even though PhO^- has a more negative charge on the nucleophilic atom; why nitrogen is the preferred nucleophile in deprotonated amides despite its less negative charge; and why soft nucleophiles attack α,β-unsaturated ketones at the softer and less positively charged β-carbon.[5] Application to hexagonal Au_7 clusters showed that the central Au atom is unusually weakly bound, with a relatively large overlap distance given its charge, rationalizing previous experimental demonstrations that doped clusters Au_6M^- (M = Ti, V, Cr, Y) prefer to replace the central Au atom with M.[5] Combined, these results show that the orbital overlap distance provides quantitative information about orbital overlap that is unavailable from electrostatics alone.

Combining the Orbital Overlap Distance and Electrostatic Potential. The present work considers the orbital overlap distance and electrostatic potential computed on molecule surfaces, an approach introduced in ref 31. Figure 2 shows a novel and biologically relevant illustration of how the electrostatic potential and the orbital overlap distance combine to distinguish coordination sites. Cysteine has three chemically distinct Lewis base sites (O, N, and S) and can form chelates in many different ways.[39] Figure 2 shows the computed structure, electrostatic potential, and orbital overlap distance of a representative conformation of deprotonated cysteine evaluated in continuum water solvent. The electrostatic potential clearly distinguishes the Lewis base sites from the surrounding molecule but does not much distinguish N from S. In contrast, the overlap distance clearly distinguishes the large orbital lobes of the chemically "soft" S Lewis base (in green) from the "harder" N Lewis base (in red). The accompanying table shows the values of the electrostatic potential and overlap distance at the points labeled with arrows.

It is important to note one caveat in interpreting these plots. The electrostatic potential and orbital overlap distance have some chemically reasonable correlation with each other. Adding electron density to a region tends to make the electrostatic potential more negative and make the overlap distance larger, as the extra electron density tends to be bound in relatively diffuse orbital lobes. However, we will endeavor to show that the orbital overlap distance includes information unavailable from the electrostatic potential alone.

In this work, we show how the overlap distance complements the electrostatic potential for representative problems in biological and medicinal chemistry. We consider the quantitative coordination chemistry of pairs of biologically relevant ions with similar charges and sizes, the different coordination sites of a metal-binding protein, and the reactivity of the rhodanine "privileged scaffold". We conclude by reanalyzing a previous medicinal chemistry study of drug candidate electrostatic potentials[12,13] and show that including the orbital overlap distance provides a nontrivial prediction for improving the in vivo activity. These results motivate adopting the overlap distance alongside electrostatic potentials in medicinal chemistry.

■ METHODS

Molecular density isosurfaces, electrostatic potentials, and the orbital overlap distance were evaluated using the open-source Multiwfn package.[23] Technical notes on computing the overlap distance using Multiwfn or the Gaussian 16 *cubegen* utility are available online at http://personal.tcu.edu/bjanesko. Electrostatic potentials were computed for isolated molecules in vacuum unless noted otherwise.[14] Ligand and protein "surfaces" were taken to be the 0.001 e/bohr^3 density isosurface.[34] This isosurface encompasses approximately 96%

of the electronic charge[40] and is considered standard in calculations of electrostatic potential.[41] Figure S3 illustrates the effect of choosing different isosurfaces. Electrostatic potentials and overlap distances are visualized using VMD version 1.9.3.[42] Visualizations of ONIOM embedded calculations show the low-level region as protein backbone ribbon diagrams colored by secondary structure.

Technical details of Figure 1 are as follows. We considered point $\mathbf{r}$ on the α-spin density isosurface defined by $\rho_\alpha(\mathbf{r}) = 0.001$ e/bohr^3. The molecular orbitals were plotted on the isosurface $|\psi(\mathbf{r})|^2 = 0.0004$ e/bohr^3, and the EDR test function was plotted on the surface where it reached 70% of its maximum value. Molecular orbitals from Hartree–Fock/aug-cc-pVQZ calculations were used.

Technical details of other figures are as follows. Input molecular orbitals were obtained from generalized Kohn–Sham DFT calculations using approximate exchange–correlation (XC) functionals[43,44] or ONIOM embedded DFT/Amber calculations,[45,46] performed using the Gaussian electronic structure package.[47] The calculations in Figures 2 and 3 used the ωB97X-D XC functional[48] and the aug-cc-pVTZ-PP basis set. Details of Figure 4 are discussed below. The calculations in Figure 5 were performed at the ωB97X-D/def2-TZVPP//ωB97X-D/6-31+G(d,p) level. The calculations in Figure 6 were performed at the B3LYP/6-31+G(d,p) level. Atom-averaged valence orbital overlap distances for the system in Figure 5 were evaluated using Hirshfeld population analysis[31,49] as described previously.[5]

The calculations in Figure 4 began from the experimental crystal structure of formylglycine-generating enzyme (PDB ID 5NXL).[50] We used the AmberTools[51] package to add hydrogen atoms and counterions, solvate with an octahedral water box, optimize the initial crystal structure coordinates, and equilibrate to room temperature using molecular dynamics simulations under isothermal–isobaric ensemble (*NPT*) conditions. The final frame of the molecular dynamics trajectory was used in ONIOM calculations with ωB97X-D/6-311G(d,p) DFT on the binding site, AMBER molecular mechanics for the remaining protein, and no treatment of surrounding water. The results in Figure 4 combine three separate ONIOM calculations, one for each binding site. Gaussian input files for all three calculations and details of the molecular dynamics simulations are provided in the Supporting Information.

The overlap distance was evaluated as shown in refs 5, 22, and 31. The overlap distance was computed from EDR($\mathbf{r}$; d_i) on an even-tempered grid of distances d_i, typically including 50 values starting at $d_{i=1}{}^{-2} = 2.5$ bohr^{-2} with an increment of 1.5 bohr^{-2}.

RESULTS AND DISCUSSION

Quantifying Coordination by Combining the Orbital Overlap Distance and Electrostatic Potential. We begin by building on previous work[5,22,29−31,33,35,37] with solid and quantitative results that the orbital overlap distance can do what the electrostatic potential cannot. We first consider the special case of systems with *nearly identical* surface electrostatic potentials and show that the surface overlap distance quantifies *remaining* differences in chemistry. Table 1 shows the differences in experimental aqueous-phase ammonia binding constants,[52] $\Delta[\log K(\mathrm{NH_3})]$, for three pairs of cations with similar sizes and charges: K^+ and Au^+, Cd^{2+} and Hg^{2+}, and Mg^{2+} and Zn^{2+}. (For example, the experimental ammonia binding constants $\log K(\mathrm{NH_3})$ are −2.8 for K^+ and 5.6 for Au^+, giving a $\Delta[\log K(\mathrm{NH_3})]$ of −8.4. Values for individual species are listed in Table S1.) These binding constants are a useful indicator of cations' relative affinities for N donors over O donors. We compare these differences in binding constant to the absolute percent differences in tabulated ionic radii[53] and the percent differences in the surface electrostatic potential and orbital overlap distance computed at those ionic radii. (For example, the tabulated ionic radius of K^+ is 138 pm, and the tabulated ionic radius of Au^+ is 137 pm, giving a percent difference $\Delta R = 100\% \times (138\ \mathrm{pm} - 137\ \mathrm{pm})/(138\ \mathrm{pm} + 137\ \mathrm{pm}) \leq 1\%$. The computed surface overlap distance of K^+ is 2.22 bohr and that of Au^+ is 1.64 bohr, giving a 15% difference. Values for individual species are listed in Table S1.)

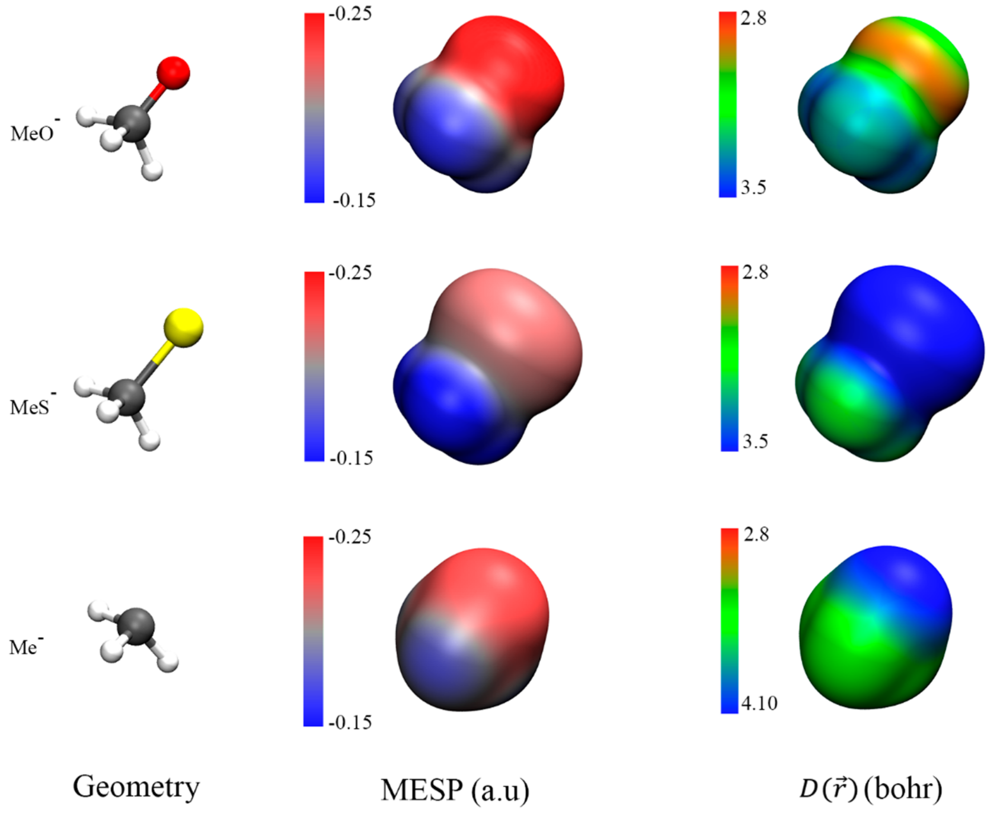
Table 1. Differences in Experimental Ammonia Binding Affinity ($\Delta[\log K(\mathrm{NH_3})]$), Tabulated Ionic Radii (ΔR), Computed Surface Electrostatic Potential (ΔV_{surf}), and Surface Overlap Distance (ΔD_{surf}) for Three Pairs of Metal Ions with Similar Ionic Radii and Electrostatic Potentials

metal ion pair	$\Delta[\log K(\mathrm{NH_3})]$	ΔR	ΔV_{surf}	ΔD_{surf}
K^+ and Au^+	−8.4	<1%	<1%	+15%
Ca^{2+} and Hg^{2+}	−9.0	1%	<1%	+13%
Mg^{2+} and Zn^{2+}	−1.9	1%	2%	+6%

Ion charge and size alone cannot explain the differences in ammonia binding, as the selected pairs of ions have sizes and electrostatic potentials that are within a few percent of each other. The additional information provided by the overlap distance clearly helps quantify the measured differences in ammonia binding affinity. In all three cases, the ion with the larger overlap distance has a weaker experimental ammonia affinity, indicating a preference for O donors. Chemically, ions with more diffuse valence orbitals and larger surface overlap distances likely have reduced Pauli repulsion from the compact (cf. Figure 2) oxygen lone pairs. Quantitatively, the difference between the ammonia binding affinities of pairs of similarly sized cations is well-approximated by the difference between the surface overlap distances as $\Delta[\log K(\mathrm{NH_3})] = -0.8\Delta D_{\mathrm{surf}}$ ($R^2 = 0.92$). Specialists should note that complete prediction of ions' aqueous-phase coordination chemistry requires careful consideration of enthalpy, entropy, solvent coordination geometry, and many other factors beyond the scope of this work. However, the overlap distance's ability to distinguish ions of similar charge and size indicates its utility beyond the electrostatic potential alone.

We next consider how the surface electrostatic potential and surface overlap distance *work together* to provide a rich picture of chemistry. Figure 3 presents the electrostatic potentials and overlap distances for three small representative Lewis bases that have *different* electrostatic potentials: MeO^-, MeS^-, and Me^-. The figure also shows the bases' gas-phase binding affinities to the hard Lewis acid H^+ and the soft Lewis acid Au^+. The electrostatic potential alone cannot capture all of the trends in binding affinities: for example, the order of electrostatic potentials is $MeS^- > Me^- > MeO^-$, but the order of H^+ binding affinities is $Me^- > MeO^- > MeS^-$. The orbital overlap distance alone *also* cannot capture all of the trends in binding affinities: for example, the order of overlap distances is $Me^- > MeS^- > MeO^-$. However, the electrostatic potential and orbital overlap distance *combined* can capture the trends in binding affinities. The electrostatic potential and overlap distance combine to distinguish the hard Lewis base

DOI: 10.1021/acs.jcim.8b00370
J. Chem. Inf. Model. 2018, 58, 1836−1846



Base	Electrostatic potential minimum V_{surf} (a.u)	Overlap distance maximum D_{surf} (bohr)	Interaction Energy (kcal/mol)	
			H^+	Au^+
MeO^-	-0.26	3.1	392	213
MeS^-	-0.21	3.7	364	220
Me^-	-0.23	4.2	427	267

Figure 3. (top) Computed structures, molecular electrostatic potentials (MESPs), and overlap distances $D(\mathbf{r})$ of three biologically relevant Lewis bases. Electrostatic potentials range from negative (blue) to positive (red) and overlap distances from compact (red) to moderate (green) to diffuse (blue). (bottom) Most negative electrostatic potential, largest overlap distance, and accurate DFT-computed gas-phase interaction energies with H^+ and Au^+ for each base.

MeO^- (large negative electrostatic potential, small D), the soft Lewis base MeS^- (less negative electrostatic potential, large D), and the very soft, very strong carbanion Lewis base Me^- (moderate negative electrostatic potential, very large D). The table in Figure 3 illustrates how the hard acid H^+ binds more tightly to MeO^- than to MeS^-, the soft acid Au^+ binds more tightly to MeS^- than to MeO^-, and both acids' tightest bond is to the strong base Me^-. This picture is consistent with Pearson's hard−soft acid−base principle that acids and bases of comparable strength prefer hard−hard and soft−soft interactions and with the fact that strong Lewis bases tend to be soft.

To further quantify this, we present an initial quantitative structure−property investigation of the data in Figure 3. We model the binding affinity of base B to acid A, ΔE(B,A) (tabulated in Figure 3), as a sum of electrostatic and orbital-overlap-distance contributions:

$$\Delta E(\mathrm{B, A}) = V_{\mathrm{surf}}(\mathrm{B})C_V(\mathrm{A}) + D_{\mathrm{surf}}(\mathrm{B})C_D(\mathrm{A}) \tag{2}$$

Fitting eq 2 to the data in Figure 3 gives $C_V(H^+) = -915$, $C_D(H^+) = 49$, $C_V(Au^+) = -234$, and $C_D(Au^+) = 49$ with a root-mean-square deviation (RMSD) in the computed binding affinities of 8 kcal/mol. The parameter values are chemically reasonable, showing that orbital overlap is relatively important for the gold cation, with $|C_V(Au^+)| < |C_V(H^+)|$. Including only electrostatic contributions in the model ($C_D(A) = 0$) increases the RMSD to 35 kcal/mol. Including only orbital-overlap-distance contributions ($C_V(A) = 0$) increases the RMSD to 31 kcal/mol. (Table S2 gives the binding affinities predicted by all three models.) Combining the electrostatic potential and

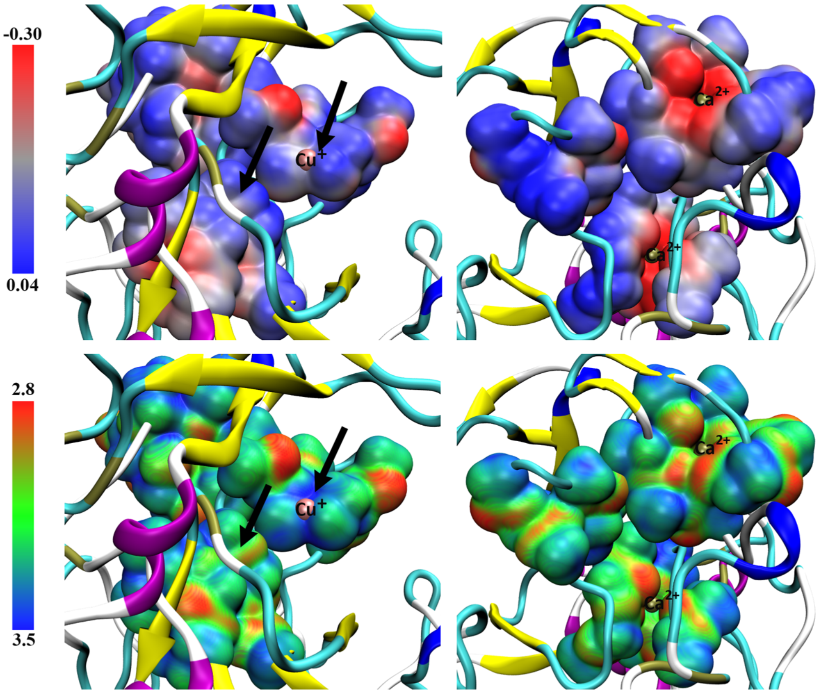


Figure 4. (top) Electrostatic potential and (bottom) overlap distance of (left) Cu^+ and (right) Ca^{2+} binding sites of FGE (PDB ID 5NXL). The electrostatic potential is most negative in the Ca^{2+} binding sites (top right, red). The overlap distance is relatively large in the cysteine sulfurs of the Cu^+ binding site (bottom left, blue). Arrows indicate representative regions with similar electrostatic potentials and different overlap distances.

orbital overlap distance gives a level of accuracy unattainable by either alone.

It is especially noteworthy that the orbital overlap distance in Figure 3 captures the "coordination" chemistry of the Me^- carbanion. Carbanion binding to H^+ of course forms a covalent C−H bond. The very large overlap distance computed for the Me^- carbanion is clearly consistent with its instability and reactivity. (Experimentally, the gas-phase pK_a of methane is substantially larger than that of methanol or methanethiol.) Electrostatics alone does *not* capture carbanion reactivity, as the electrostatic potential of MeO^- is more negative than that of Me^-. Figure 3 thus confirms our previous demonstrations[5,22,29−31,33,35,37] that the overlap distance captures orbital-dependent aspects of chemical reactivity that are different from and complementary to the aspects captured by the electrostatic potential.

Metal Binding Sites. We next show how the electrostatic potential and overlap distance combine to distinguish chemically distinct metal binding sites on a single protein. We choose formylglycine-generating enzyme (FGE),[54−59] a representative copper-dependent[60] protein catalyzing C−H activation.[56,60] FGE has one high-affinity[56] binding site for Cu^+[61] and two binding sites for Ca^{2+}.[56] Figure 4 shows the electrostatic potential and overlap distance in *Thermomonospora curvata* FGE computed from the experimental crystal structure of its complex with Ag^+, a redox-stable Cu^+ mimic (PDB ID 5NXL).[50] The Cu^+ binding site contains cysteines Cys269 and Cys274. Ca^{2+} binding site Ca1 contains the carboxylate group of Asp202 and the hydroxyl groups of Ile189, Asn188, and Tyr204 in addition to a coordinated water molecule. Site Ca2 contains the hydroxyl groups of Asn222, Val223, Gly225, and Val227.

Figure 4 shows two views (left and right) of the computed electrostatic potential (top) and overlap distance (bottom) of FGE. The enzyme has a high-affinity binding site for Cu^+ visible in the left view, and two Ca^{2+} binding sites are visible in the right view. The two Ca^{2+} binding sites have a high concentration of negative charge from the carboxylate and hydroxyl groups that bind Ca^{2+}. This high concentration of negative charge is visible as bright-red regions in the plot of electrostatic potential (top right), clearly distinct from the other regions of the enzyme surface. Given only the computed electrostatic potentials in Figure 4, it would be straightforward to determine the locations of the two Ca^{2+} binding sites.

The situation is different for the high-affinity Cu^+ binding site. The key cysteine residues do not give as high a concentration of negative charge. The Cu^+ binding site is not visible as a bright-red region in the plot of electrostatic potential. It is (at the present level of theory) a gray-blue region of modest negative charge, similar to many other gray-blue regions on the enzyme surface (see, e.g., the two black arrows). Given only the computed electrostatic potentials in Figure 4, it would not be straightforward to determine the location of the high-affinity Cu^+ binding site.

The orbital overlap distance provides the necessary additional information. The orbital overlap distance shows the chemically "soft" cysteine Lewis bases in the high-affinity Cu^+ binding site as a bright-blue region (bottom left) and the "harder" Ca^{2+} binding sites as bright-red regions (bottom right). This distinguishes the region that binds copper. Given both the electrostatic potential and overlap distance surfaces in Figure 4, it would be more straightforward to determine the locations of all three binding sites. While this example is arguably artificial, as practicing chemists do not interpret computed electrostatic potentials in isolation, it does highlight how the overlap distance and electrostatic potential combine to provide a richer picture of binding site chemistry.

Similar results are seen for other binding sites. Figures S2 and S3 show the computed electrostatic potential and overlap distance for the cysteine dimer in the active site of the gold-

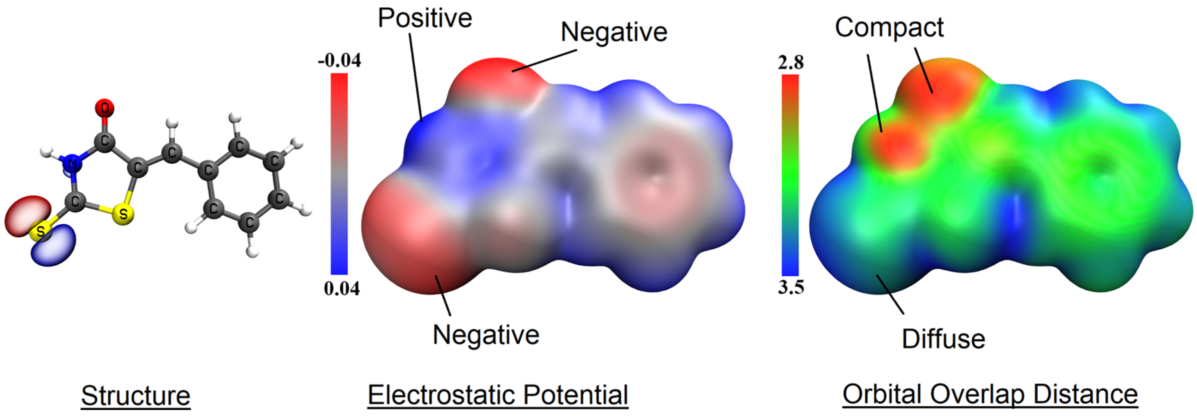


Figure 5. (left) Structure and HOMO−1, (middle) surface electrostatic potential, and (right) surface overlap distance of benzylidene-substituted rhodanine.

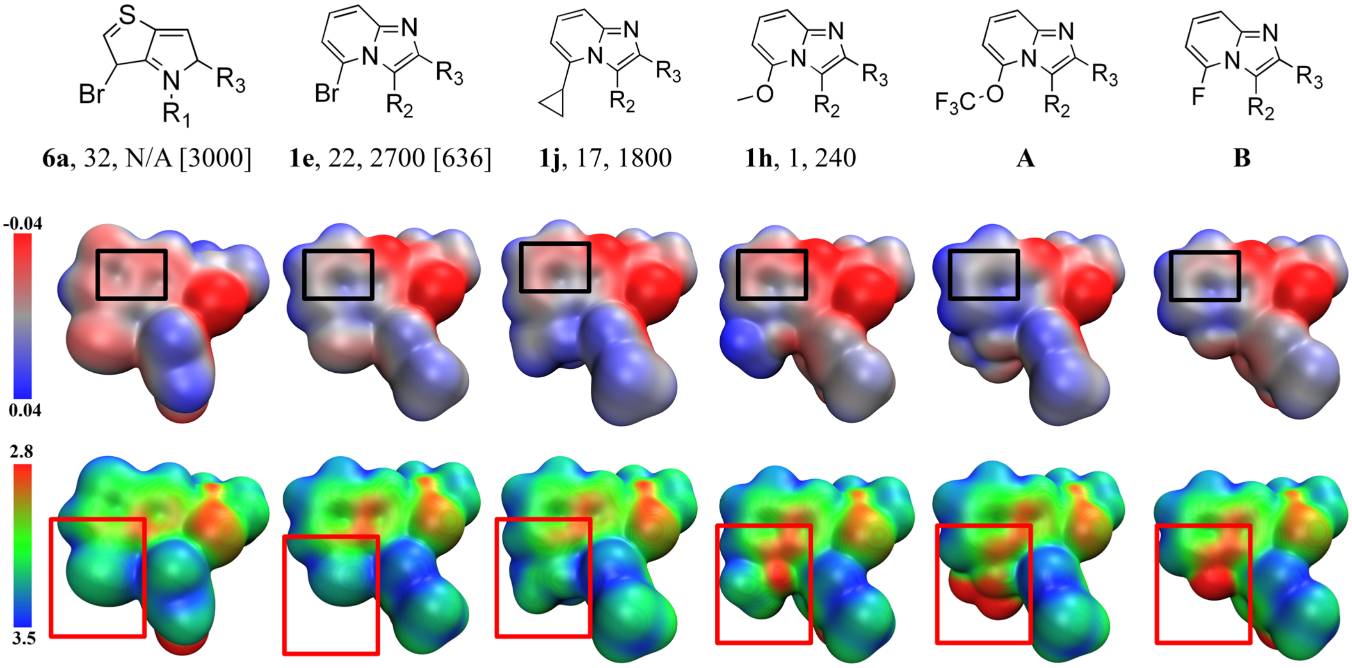


Figure 6. (top) Structures, in vitro CENP-E IC_{50} values (in nM), and in vivo HeLa cell proliferation values (in nM) [p-HH3 EC_{50} (nM)] of CENP-E inhibitors **6a**, **1e**, **1j**, and **1h** from refs 12 and 13 and inhibitors **A** and **B** proposed here. R_1 = *p*-fluorobenzene; R_2 = *m*-methyl-*p*-fluorobenzene; R_3 = $C(=O)N(C_2H_4NMe_2)PhCl_2$. (middle and bottom) Predicted electrostatic potentials and overlap distances of the fused-ring regions. The neutral fused-ring electrostatic potential (black boxes) was previously shown to be correlated with the in vitro activity.[12] We propose that small substituent orbital overlap distances (red boxes) are correlated with the in vivo activity.

binding protein GolB. The dimer has a large overlap distance, consistent with the relative chemical softness. Specialists should note that calculations in four different basis sets, including or omitting a continuum model for the water solvent, and using three different choices of density isosurface, all give qualitatively similar results.

Promiscuous Binding. The electrostatic potential and orbital overlap distance can contribute to ongoing discussions in the medicinal chemistry literature. Rhodanines have been identified as problematic "promiscuous hitters" that interact nonspecifically with many targets.[62,63] However, other studies suggest that rhodanines are "privileged scaffolds" that are useful in drug design.[11,64] A recent experimental and computational study suggested that rhodanines' HOMO and negative electrostatic potential are strongly localized at the exocyclic sulfur and proposed that this was responsible for rhodanine's distinct intermolecular interaction profile.[11]

Figure 5 examines the orbitals, electrostatic potential, and orbital overlap distance of the benzylidene-substituted rhodanine derivative studied in ref 11. We find that the electrostatic potential minimum occurs on the exocyclic oxygen, not on the less electronegative exocyclic sulfur. In contrast, the overlap distance is largest on the exocyclic sulfur, consistent with the HOMO−1. The orbital overlap distance in Figure 5 is relatively large (blue and green) on the exocyclic sulfur, distinguishing it from the more compact exocyclic oxygen (red) and aromatic sulfur (green). Quantitatively, Hirshfeld population analysis[31,49] shows that the exocyclic sulfur has an atomic charge of −0.22*e* and an average valence overlap distance of 1.94 bohr, making it more charged and more diffuse than the ring sulfur (+0.04*e*, 1.85 bohr) but less charged and much more diffuse than the exocyclic oxygen (−0.28*e*, 1.32 bohr). Chemically, the electrostatic potential and orbital overlap distance are reminiscent of deprotonated

DOI: 10.1021/acs.jcim.8b00370
J. Chem. Inf. Model. 2018, 58, 1836−1846

amides. Deprotonated amides perform nucleophilic attack by the less negative nitrogen atom rather than the more negative oxygen atom, consistent with the nitrogen atom's substantially larger orbital overlap distance.[5] We suggest that rhodanine's promiscuous binding arises from the less negative exocyclic sulfur rather than the more negative exocyclic oxygen, consistent with the former's substantially larger orbital overlap distance.

Extending Frontier Orbital Analysis. Rhodanine also illustrates how the orbital overlap distance extends classical Fukui analyses of individual frontier orbitals. The calculations in Figure 5 used the modern ωB97X-D DFT exchange−correlation functional. These calculations predict that the rhodanine derivative's HOMO is not strongly localized at the exocyclic sulfur but is instead delocalized over the entire π system. (Figure S3 shows the rhodanine HOMO, HOMO−1, electrostatic potential, and overlap distance evaluated with different methods.) The localized orbital in Figure 5 is the ωB97X-D-computed HOMO−1. Calculations with the older BLYP[65,66] XC functional used in ref 11 invert the order of the HOMO and HOMO−1, matching the results from that reference. In contrast, Figure S4 shows that ωB97X-D, BLYP, and ωB97X-D continuum solvent calculations give *nearly indistinguishable* electrostatic potential and overlap distance plots. The robust nature of the overlap distance further illustrates its value as an interpretive tool.

CENP-E Inhibitor Design. We conclude by combining the electrostatic potential and overlap distance to make a nontrivial prediction in structure-based drug design. We revisit a published "real-world" application of ligand electrostatic potentials in a combined experimental and computational study. We show that combining the computed ligand electrostatic potentials with new calculations of the same ligands' overlap distances provides a prediction for a previously untested inhibitor.

We consider a series of studies by Hirayama and co-workers developing centromere-associated protein E (CENP-E) inhibitors.[12,13] CENP-E is a mitotic spindle motor protein and a promising target for cancer therapies. A combination of high-throughput screening, structure−activity relationship measurements, and homology model docking identified lead compound **6a** (Figure 6). The authors of those studies performed SAR analysis by inspecting computed electrostatic potential maps. On the basis of these inspections, the authors suggested that the in vitro activity was correlated with a *neutral* electrostatic potential on the aromatic ring moiety, the region highlighted with black boxes in Figure 6. However, compound **6a** possessed insufficient *cellular* activity despite its neutral aromatic ring. This motivated the synthesis of the new lead compound **1e**, whose neutral aromatic region was combined with higher in vivo activity. Subsequent derivatization produced improved species **1j** and **1h**, leading to the eventual identification of a potent 5-methoxyimidazo[1,2-*a*]pyridine derivative.

Before continuing, we note that solubility, entropic and enthalpic effects, and an enormous number of other factors affect inhibitor in vitro and in vivo activities. These effects are far beyond the scope of this work. However, we suggest that the authors of refs 12 and 13 successfully used computed inhibitor electrostatic potentials to optimize one aspect of ligand in vitro activity. We predict that computed inhibitor overlap distances allow us to optimize an aspect of ligand in vivo activity.

Figure 6 shows the structures, measured in vivo and in vitro activities, and computed electrostatic potentials and overlap distances of **6a**, **1e**, **1j**, and **1h** (numbering follows the experimental references). The figure also shows the computed structures, electrostatic potentials, and overlap distances of the new compounds **A** and **B** proposed in this work. The calculations truncated the dichlorobenzyl moiety, following ref 12.

The black boxes in Figure 6 show the electrostatic potentials in the region of the aromatic ring moiety. In ref 12 it was found that compounds with high in vitro activity have a neutral electrostatic potential in this region. Our computed electrostatic potentials agree with this result: all of the ligands in Figure 6 have a neutral electrostatic potential in the region in the black box.

The red boxes in Figure 6 show the orbital overlap distances in the region of the aromatic ring substituents. The orbital overlap distance in this region is relatively large (blue-green) for compounds **6a** and **1e**, smaller (green) for compound **1a**, and smaller yet (red and green) for compound **1b**. Compounds with high in vivo activity appear to have a compact orbital overlap distance in this region. The orbital overlap distance provides a new and nontrivial prediction: *compact substituents with small overlap distance in the aromatic ring region improve in vivo activity*.

On the basis of this analysis, we predict that novel compound **A**, obtained by replacing the methoxy group with a trifluoromethyl ether (a strategy increasingly adopted in medicinal chemistry and drug discovery[67]), will exhibit improved in vivo activity. We thank an anonymous reviewer for suggesting compound **B**, obtained by replacing the methoxy group with a fluoro substituent. To our knowledge, compounds **A** and **B** have not been previously proposed as CENP-E inhibitors. However, Figure 6 shows that both A and B combine a relatively neutral electrostatic potential in the aromatic ring moiety with a compact orbital overlap distance in the region of the aromatic ring substituents. On this basis, we predict that these compounds will exhibit higher activities than **1h**. We predict that replacing the corresponding methoxy substituent with a fluorine or a trifluoromethyl ether in the final imidazo[1,2-*a*]pyridine derivative, (+)-**12** in ref 13, could provide additional increases in activity. Figure S5 shows the full chemical structures of (+)-**12** and the proposed trifluoromethylated CENP-E inhibitor based on **A**.

CONCLUSIONS

The examples presented above show how the orbital overlap distance can complement electrostatic potential maps in biological and medicinal chemistry, providing a clearer picture of orbital overlap effects in large biochemical systems. The orbital overlap distance distinguishes which parts of a biomolecule surface come from diffuse, polarizable orbitals. The orbital overlap distance quantifies the different experimental coordination chemistries of pairs of ions with similar sizes and charges, quantifies hard−soft acid−base aspects of coordination chemistry, helps distinguish the binding sites of soft Cu^+ and hard Ca^{2+} cations on formylglycine-generating enzyme, highlights the orbital-driven aspects of rhodanine's promiscuous binding, and provides new nontrivial predictions in structure-based drug design. These results motivate using the overlap distance alongside electrostatic potential maps in interpreting MO calculations in medicinal chemistry.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00370.

Computational details, supplementary tables and figures, Cartesian coordinates, and Gaussian input files (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: b.janesko@tcu.edu.

ORCID

Peng Tao: 0000-0002-2488-0239

Benjamin G. Janesko: 0000-0002-2572-5273

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (DMR-1505343).

■ ABBREVIATIONS

DFT, density functional theory; ELF, electron localization function; QSAR, quantitative structure−activity relationship; CENP-E, centromere-associated protein E; FGE, formylglycine-generating enzyme; HOMO, highest occupied molecular orbital

■ REFERENCES

(1) Pauling, L. The Nature Of the Chemical Bond. Application of Results Obtained From the Quantum Mechanics and From a Theory of Paramagnetic Susceptibility to the Structures of Molecules. *J. Am. Chem. Soc.* **1931**, *53* (4), 1367−1400.

(2) Fukui, K. Role of Frontier Orbitals in Chemical Reactions. *Science* **1982**, *218* (4574), 747−754.

(3) Hoffmann, R. Building Bridges Between Inorganic and Organic Chemistry (Nobel Lecture). *Angew. Chem., Int. Ed. Engl.* **1982**, *21* (10), 711−724.

(4) Klopman, G. Chemical reactivity and the concept of charge- and frontier-controlled reactions. *J. Am. Chem. Soc.* **1968**, *90* (2), 223−234.

(5) Mehmood, A.; Janesko, B. G. An Orbital-Overlap Complement to Atomic Partial Charge. *Angew. Chem., Int. Ed.* **2017**, *56* (24), 6878−6881.

(6) Prozialeck, W. C.; Lamar, P. C. Interaction of Cadmium (Cd2+) With a 13-Residue Polypeptide Analog of a Putative Calcium-Binding Motif of E-Cadherin. *Biochim. Biophys. Acta, Mol. Cell Res.* **1999**, *1451* (1), 93−100.

(7) Varbanov, H. P.; Jakupec, M. A.; Roller, A.; Jensen, F.; Galanski, M.; Keppler, B. K. Theoretical Investigations and Density Functional Theory Based Quantitative Structure−Activity Relationships Model for Novel Cytotoxic Platinum(IV) Complexes. *J. Med. Chem.* **2013**, *56* (1), 330−344.

(8) DeChancie, J.; Houk, K. N. The Origins of Femtomolar Protein-Ligand Binding: Hydrogen-Bond Cooperativity and Desolvation Energetics in the Biotin-(Strept)Avidin Binding Site. *J. Am. Chem. Soc.* **2007**, *129* (17), 5419−5429.

(9) Friesner, R. A.; Guallar, V. Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.

(10) Akimov, A. V.; Prezhdo, O. V. Large-Scale Computations in Chemistry: A Bird's Eye View of a Vibrant Field. *Chem. Rev.* **2015**, *115* (12), 5797−5890.

(11) Mendgen, T.; Steuer, C.; Klein, C. D. Privileged Scaffolds or Promiscuous Binders: A Comparative Study on Rhodanines and Related Heterocycles in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55* (2), 743−753.

(12) Hirayama, T.; Okaniwa, M.; Imada, T.; Ohashi, A.; Ohori, M.; Iwai, K.; Mori, K.; Kawamoto, T.; Yokota, A.; Tanaka, T.; Ishikawa, T. Synthetic studies of centromere-associated protein-E (CENP-E) inhibitors: 1. Exploration of fused bicyclic core scaffolds using electrostatic potential map. *Bioorg. Med. Chem.* **2013**, *21* (17), 5488−5502.

(13) Hirayama, T.; Okaniwa, M.; Banno, H.; Kakei, H.; Ohashi, A.; Iwai, K.; Ohori, M.; Mori, K.; Gotou, M.; Kawamoto, T.; Yokota, A.; Ishikawa, T. Synthetic Studies on Centromere-Associated Protein-E (CENP-E) Inhibitors: 2. Application of Electrostatic Potential Map (EPM) and Structure-Based Modeling to Imidazo[1,2-*a*]pyridine Derivatives as Anti-Tumor Agents. *J. Med. Chem.* **2015**, *58* (20), 8036−8053.

(14) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268* (5214), 1144−1149.

(15) Zhu, X.; Yu, W.; McBride, R.; Li, Y.; Chen, L.-M.; Donis, R. O.; Tong, S.; Paulson, J. C.; Wilson, I. A. Hemagglutinin homologue from H17N10 bat influenza virus exhibits divergent receptor-binding and pH-dependent fusion activities. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (4), 1458−1463.

(16) Jagusch, C.; Negri, M.; Hille, U. E.; Hu, Q.; Bartels, M.; Jahn-Hoffmann, K.; Mendieta, M. A. E. P.-B.; Rodenwaldt, B.; Müller-Vieira, U.; Schmidt, D.; Lauterbach, T.; Recanatini, M.; Cavalli, A.; Hartmann, R. W. Synthesis, biological evaluation and molecular modelling studies of methyleneimidazole substituted biaryls as inhibitors of human 17α-hydroxylase-17,20-lyase (CYP17). Part I: Heterocyclic modifications of the core structure. *Bioorg. Med. Chem.* **2008**, *16* (4), 1992−2010.

(17) Van Damme, S.; Bultinck, P. 3D QSAR based on conceptual DFT molecular fields: Antituberculotic activity. *J. Mol. Struct.: THEOCHEM* **2010**, *943* (1−3), 83−89.

(18) Faver, J.; Merz, K. M. Utility of the Hard/Soft Acid-Base Principle via the Fukui Function in Biological Systems. *J. Chem. Theory Comput.* **2010**, *6* (2), 548−559.

(19) Galanakis, D.; Davis, C. A.; Ganellin, C. R.; Dunn, P. M. Synthesis and Quantitative Structure-Activity Relationship of a Novel Series of Small Conductance Ca2+-Activated K+ Channel Blockers Related to Dequalinium. *J. Med. Chem.* **1996**, *39* (2), 359−370.

(20) Bultinck, P.; Carbó-Dorca, R. Molecular quantum similarity using conceptual DFT descriptors. *Proc. - Indian Acad. Sci., Chem. Sci.* **2005**, *117* (5), 425−435.

(21) Roos, G.; Geerlings, P.; Messens, J. Enzymatic Catalysis: The Emerging Role of Conceptual Density Functional Theory. *J. Phys. Chem. B* **2009**, *113* (41), 13465−13475.

(22) Janesko, B. G.; Scalmani, G.; Frisch, M. J. How far do electrons delocalize? *J. Chem. Phys.* **2014**, *141* (14), 144104.

(23) Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33* (5), 580−592.

(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision A.03; Gaussian, Inc.: Wallingford, CT, 2016.

(25) Becke, A. D.; Edgecombe, K. E. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.* **1990**, *92* (9), 5397−5403.

(26) Mulliken, R. S. Electronic Population Analysis on LCAO−MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23* (10), 1833−1840.

(27) Wiberg, K. B. Application of the Pople-Santry-Segal CNDO method to the cyclopropylcarbinyl and cyclobutyl cation and to bicyclobutane. *Tetrahedron* **1968**, *24* (3), 1083−1096.

(28) Williams, C. F.; Herbert, J. M. Influence of Structure on Electron Correlation Effects and Electron-Water Dispersion Interactions in Anionic Water Clusters. *J. Phys. Chem. A* **2008**, *112* (27), 6171−6178.

(29) Janesko, B. G.; Scalmani, G.; Frisch, M. J. Quantifying solvated electrons' delocalization. *Phys. Chem. Chem. Phys.* **2015**, *17* (28), 18305−18317.

(30) Janesko, B. G.; Jones, S. I. Quantifying the delocalization of surface and bulk F-centers. *Surf. Sci.* **2017**, *659*, 9−15.

(31) Janesko, B. G.; Wiberg, K. B.; Scalmani, G.; Frisch, M. J. Electron Delocalization Range in Atoms and on Molecular Surfaces. *J. Chem. Theory Comput.* **2016**, *12* (7), 3185−3194.

(32) Miao, M.-S.; Hoffmann, R. High-Pressure Electrides: The Chemical Nature of Interstitial Quasiatoms. *J. Am. Chem. Soc.* **2015**, *137* (10), 3631−3637.

(33) Janesko, B. G.; Scalmani, G.; Frisch, M. J. Quantifying Electron Delocalization in Electrides. *J. Chem. Theory Comput.* **2016**, *12*, 79−91.

(34) Bader, R. F. W.; Henneker, W. H.; Cade, P. E. Molecular Charge Distributions and Chemical Binding. *J. Chem. Phys.* **1967**, *46* (9), 3341−3363.

(35) Janesko, B. G. Topological analysis of the electron delocalization range. *J. Comput. Chem.* **2016**, *37* (21), 1993−2005.

(36) Savin, A.; Silvi, B.; Colonna, F. Topological analysis of the electron localization function applied to delocalized bonds. *Can. J. Chem.* **1996**, *74* (6), 1088−1096.

(37) Mehmood, A.; Janesko, B. G. The electron delocalization range in stretched bonds. *Int. J. Quantum Chem.* **2016**, *116* (23), 1783−1795.

(38) Deng, Z.; Martyna, G. J.; Klein, M. L. Electronic states in metal-ammonia solutions. *Phys. Rev. Lett.* **1993**, *71* (2), 267−270.

(39) Norbury, A. H.; Sinha, A. I. P. The co-ordination of ambidentate ligands. *Q. Rev., Chem. Soc.* **1970**, *24* (1), 69−94.

(40) Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987**, *109* (26), 7968−7979.

(41) Kolář, M. H.; Hobza, P. Computer Modeling of Halogen Bonds and Other σ-Hole Interactions. *Chem. Rev.* **2016**, *116* (9), 5155−5187.

(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.

(43) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(44) Kohn, W.; Sham, L. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(45) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(t-Bu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100* (50), 19357−19363.

(46) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179−5197.

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian Development Version*, revision I.11; Gaussian, Inc.: Wallingford, CT, 2016.

(48) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615−6620.

(49) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chim. Acta (Berl.)* **1977**, *44*, 129−138.

(50) Ajees, A. A.; Marapakala, K.; Packianathan, C.; Sankaran, B.; Rosen, B. P. Structure of an As(III) *S*-adenosylmethionine methyltransferase: insights into the mechanism of arsenic biotransformation. *Biochemistry* **2012**, *51* (27), 5476−5485.

(51) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R.C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2017*; University of California: San Francisco, 2017.

(52) Hancock, R. D.; Martell, A. E. Ligand design for selective complexation of metal ions in aqueous solution. *Chem. Rev.* **1989**, *89* (8), 1875−1914.

(53) Shannon, R. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32* (5), 751−767.

(54) Appel, M. J.; Bertozzi, C. R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* **2015**, *10* (1), 72−84.

(55) Hudak, J. E.; Barfield, R. M.; de Hart, G. W.; Grob, P.; Nogales, E.; Bertozzi, C. R.; Rabuka, D. Synthesis of Heterobifunctional Protein Fusions Using Copper-Free Click Chemistry and the Aldehyde Tag. *Angew. Chem., Int. Ed.* **2012**, *51* (17), 4161−4165.

(56) Meury, M.; Knop, M.; Seebeck, F. P. Structural Basis for Copper−Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew. Chem., Int. Ed.* **2017**, *56* (28), 8115−8119.

(57) Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L.-X.; Bertozzi, C. R. Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjugate Chem.* **2014**, *25* (4), 788−795.

(58) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **2007**, *3*, 321.

(59) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130* (37), 12240−12241.

(60) Knop, M.; Engi, P.; Lemnaru, R.; Seebeck, F. P. In Vitro Reconstitution of Formylglycine-Generating Enzymes Requires Copper(I). *ChemBioChem* **2015**, *16* (15), 2147−2150.

(61) Knop, M.; Dang, T. Q.; Jeschke, G.; Seebeck, F. P. Copper is a Cofactor of the Formylglycine-Generating Enzyme. *ChemBioChem* **2017**, *18* (2), 161−165.

(62) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from

Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719−2740.

(63) Tomašić, T.; Peterlin Mašič, L. Rhodanine as a scaffold in drug discovery: a critical review of its biological activities and mechanisms of target modulation. *Expert Opin. Drug Discovery* **2012**, *7* (7), 549−560.

(64) Nitsche, C.; Schreier, V. N.; Behnam, M. A. M.; Kumar, A.; Bartenschlager, R.; Klein, C. D. Thiazolidinone−Peptide Hybrids as Dengue Virus Protease Inhibitors with Antiviral Activity in Cell Culture. *J. Med. Chem.* **2013**, *56* (21), 8389−8403.

(65) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(66) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37* (2), 785−789.

(67) Landelle, G.; Panossian, A.; Leroux, F. R. Trifluoromethyl Ethers and −Thioethers as Tools for Medicinal Chemistry and Drug Discovery. *Curr. Top. Med. Chem.* **2014**, *14* (7), 941−951.