JCTC Journal of Chemical Theory and Computation

Continuous Localized Orbital Corrections to Density Functional Theory: B3LYP-CLOC

Michelle Lynn Hall, Jing Zhang, Arteum D. Bochevarov, and Richard A. Friesner*

Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States

Received July 26, 2010

Abstract: Our previous works have demonstrated the ability of our localized orbital correction (LOC) methodology to greatly improve the accuracy of various thermochemical properties at the stationary points of the density functional theory (DFT) reaction coordinate (RC). Herein, we extend this methodology from stationary points to the entire RC connecting any stationary points by developing continuous localized orbital corrections (CLOCs). We show that the resultant method, DFT-CLOC, is capable of producing RCs with far greater accuracy than uncorrected DFT and yet requires negligible computational cost beyond the uncorrected DFT calculations. Various post-Hartree−Fock (post-HF) reaction coordinate profiles were used, including a sigmatropic shift, Diels−Alder reaction, electrocyclization, carbon radical, and three hydrogen radical reactions to show that this method is robust across multiple reaction types of general interest.

I. Introduction

Density functional theory (DFT)[1] has proven a very useful theoretical tool for computing atomic and molecular electronic structures. In comparison to post-Hartree−Fock methods, DFT methods are capable of calculating relatively large systems and transition-metal-containing systems and, therefore, are widely used in quantum chemistry and condensed matter physics. The accuracy of DFT methods is essentially dependent on the density functional used, which is always an approximation of the hypothetical exact density functional. During past decades, many attempts to construct a more accurate functional have been undertaken, starting from either first principles or empirical fitting, or both.[2] However, the approximate nature of extant density functionals inevitably weakens the robustness of DFT performance in predicting, in particular, thermodynamic properties.[3]

Two distinct methods can be envisioned to tackle some of the problems that still continue to plague DFT: (a) Many researchers have had remarkable success by developing wholly new density functionals.[4] (b) Alternatively, one can envision creating a new functional by simply taking an existing functional and adding terms on top of it. These terms can be used to target systematic errors endemic to each functional.

In previous publications, we have shown that the accuracy of DFT can be greatly improved for various thermochemical properties with the use of localized orbital corrections, or LOCs.[5] These LOCs have been developed to treat stationary points (i.e., reactants, products, and transition states) and are based on a chemically intuitive dissection of each stationary point's electronic structure into valence bond terms. Further, because the LOCs are applied *a posteriori* using a simple noniterative computational algorithm, they require negligible computational cost beyond standard DFT calculations. With the application of LOCs, atomization energies, ionization potentials, electron affinities, enthalpies of reaction, and barrier heights can all be obtained with very good accuracy for stationary points.

In this work, we extend our methodology beyond the treatment of discrete stationary points with the goal of providing energetics for the entire reaction coordinate (RC) of a chemical reaction. As we already have developed LOCs to treat the reactant, transition state, and product for an arbitrary reaction, an obvious next step is to interpolate these LOCs for all intermediate points and develop what we shall refer to as continuous localized orbital corrections, CLOCs.

* rich@chem.columbia.edu.

10.1021/ct100418n
Published on Web 11/08/2010

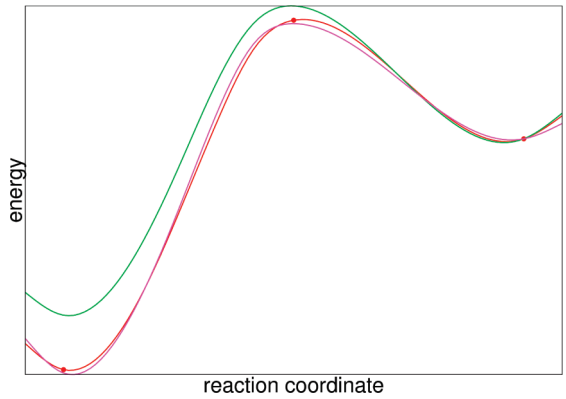


Figure 1. Reaction coordinate for an arbitrary reaction, where B3LYP (green), B3LYP-LOC (red dots), and B3LYP-CLOC (red line) are compared to an accurate benchmark (pink).

This is depicted schematically in Figure 1, where the B3LYP-LOC stationary point energies are shown (red points), connected with the B3LYP-CLOC energy curve (also red). While previous publications defined B3LYP-LOC energies only at the stationary points (red points), B3LYP-CLOC energies are defined all along the reaction coordinate (red curve). The latter is the subject of the present work.

A naming convention implicit from the previous discussion is that we use "LOC" to describe the discrete corrections, those at the stationary points, exclusively. "CLOC" is used to describe the continuous corrections, for all points that are not stationary points. This is represented schematically in Figure 1, where the B3LYP-LOC energies are defined at the stationary points (red dots), whereas the B3LYP-CLOC energies are defined all along the reaction coordinate (red line).

In order for our model to be consistent, any CLOC computed at a stationary point must agree with the LOC for that same stationary point. Importantly, it is not possible to compute a LOC for a point other than a stationary point because LOC parameters have been developed for the stationary points only. To treat points other than stationary points, interpolation is necessary in a continuous fashion, hence the necessity of the CLOC method.

It should be emphasized that LOCs are simple numerical corrections that should improve the accuracy of the DFT-predicted electronic energy. LOCs are added to the DFT energy *a posteriori* and therefore cannot be used to improve DFT-predicted geometries of molecular systems. In order to perform geometry optimizations, and hence have our corrections affect the geometry (at least in theory) and not just the energy, we must be able to calculate gradients of these corrections. In order to be able to perform geometry optimizations, with CLOCs having an effect on the changes in geometry, their first derivatives with respect to nuclear displacements must be computed. This necessitates the extension of these discrete LOCs into a continuous form that connects the various stationary points. The innovations described in this work extend LOCs to not only treat nonstationary points but also to contribute to the optimization of molecular geometries.

Herein, we present the results of our CLOC development and show that it can be applied to seven reaction coordinate profiles with greatly improved results compared to uncorrected B3LYP. We also provide comparison with the M06-2X[6] functional, which has a substantially improved performance for reaction energetics as compared to B3LYP. Note that while we have examined points along the reaction profile exclusively in this paper, this is not a necessary condition for the application of CLOCs. Points off the reaction coordinate can also be treated, as described in more detail in section III.A.

In this publication, as in others, we have focused our efforts on corrections to the well-established B3LYP functional. At the same time, we have previously tested our LOC methodology in combination with other important functionals including the M05-2X and M06-2X[6] functionals of Truhlar and co-workers.[5c] We find that no functional tested to date combines with the LOC method as favorably as B3LYP. Nevertheless, it is still possible for other functionals not yet tested to produce more accurate results in combination with LOCs than B3LYP-LOC itself.

II. Overview of the B3LYP-LOC Methodology

The B3LYP-LOC model has been successfully employed to reduce errors endemic to DFT across a wide range of thermodynamic properties including atomization energies,[5a] ionization potentials and electron affinities,[5b] enthalpies of reaction,[5c,e] and barrier heights,[5e] as shown in Table 1 and Figure 2 below.

We have previously asserted that this impressive reduction in error upon application of LOCs is not fortuitous but rather reflects the systematic nature of the errors that are intrinsic to DFT in general and specialized to the specific errors characteristic of B3LYP. The LOCs dramatically reduce such errors by assigning fitting parameters dependent upon the

Table 1. Performance of B3LYP vs B3LYP-LOC for Various Chemical Properties

	mean unsigned error (MUE) (kcal/mol)		number of parameters (LOCs) employed[a]	size of data set	ref
	B3LYP	B3LYP-LOC			
atomization energies	4.8	0.8	22 (22)	222	5a
ionization potentials and electron affinities	3.2	0.9	45 (23)	134	5b
enthalpy of reactions	4.9	0.9	28 (0)	139	5c, 5e
barrier heights	3.2	1.2	36 (8)	105	5e

[a] The number of new parameters developed in each work is shown in parentheses. Specifically, the same 22 parameters developed initially for atomization energies are used for all other calculations: ionization potentials and electron affinities, enthalpies of reactions, and barrier heights. An additional 23, 0, and 8 parameters are developed specifically for these calculations, respectively. Some of the 23 parameters developed uniquely for ionization potentials and electron affinities were applied to enthalpies of reaction and barrier heights of ionic reactions.

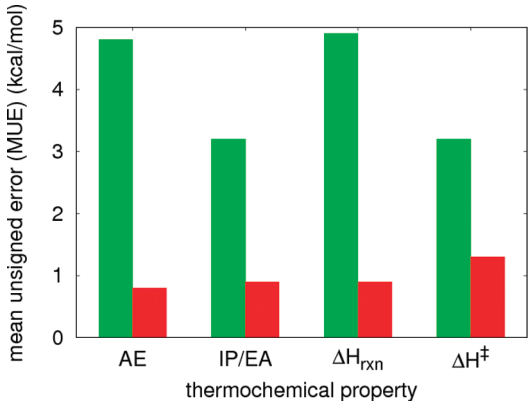


Figure 2. Performance of B3LYP (green) vs B3LYP-LOC (red) for various thermochemical properties including atomization energies (AE), ionization potentials and electron affinities (IP/EA), enthalpies of reaction (ΔH_{rxn}), and barrier heights ($\Delta H^{\ddagger}$). Data shown are taken from publications referenced in Table 1.

local environment of an electron pair or single electron. These parameters are then assumed to be transferable across molecular species. As is discussed in detail in ref 5, the dominant error in B3LYP can be identified as a difficulty in accurately modeling variations in nondynamical electron correlation across different types of chemical bonds, lone pairs, hybridization states, chemical environments, and singly vs doubly occupied orbitals.[7] The LOCs yield a more accurate representation of this variation as a function of local chemical environment.

Assigning LOCs to a particular molecular system is often straightforward. On the basis of the atomic coordinates of the molecule, a valence bond structure can be proposed. Some of the characteristics of valence bond structures have been identified as contributing to DFT's systematic errors. Accordingly, each of these particular characteristics is assigned a LOC to mitigate its error. In previous publications, all LOC values, C_k, were determined using linear regression such that they minimize the deviation between B3LYP and the reference value for many different thermochemical properties computed with several large data sets, as summarized in Table 1.[5] (A complete list of all of the LOCs and their values, C_k, is provided in the Supporting Information.) The total LOC(x) for any system x is then given simply by the sum of all individual LOCs' optimally determined values, C_k, multiplied by their number of occurrences, N_k, i.e., the number of times the particular valence bond characteristic associated with them is counted.

$$\mathrm{LOC}(x) = \sum_k N_k C_k \qquad (1)$$

These LOCs are then used in a straightforward manner to correct the enthalpy of reaction, for example. The expression for B3LYP reaction enthalpy is the difference in the enthalpies of products and reactants:

$$\Delta H_{rxn}^{B3LYP} = \sum_{products} \Delta H_{products}^{B3LYP} - \sum_{reactants} \Delta H_{reactants}^{B3LYP} \qquad (2)$$

Analogously, B3LYP-LOC reaction enthalpy is given by the LOC-corrected enthalpy differences in products and reactants:

$$\Delta H_{rxn}^{B3LYP-LOC} = \sum_{products} \Delta H_{products}^{B3LYP-LOC} - \sum_{reactants} \Delta H_{reactants}^{B3LYP-LOC} \qquad (3)$$

or equivalently,

$$\Delta H_{rxn}^{B3LYP-LOC} = \Delta H_{rxn}^{B3LYP} + \Delta \mathrm{LOC}_{rxn} \qquad (4)$$

where $\Delta \mathrm{LOC}_{rxn}$ is defined as the difference between LOC(products) and LOC(reactants). For example, consider the reaction in Scheme 1. Each species involved is assigned characteristics summarized in Table 2.

The B3LYP-LOC reaction enthalpy may be written as

$$\Delta H_{rxn}^{B3LYP-LOC}(CH_3 + CH_2{=}CH_2 \rightarrow CH_3{-}CH_2{-}CH_2) = \Delta H_{rxn}^{B3LYP}(CH_3 + CH_2{=}CH_2 \rightarrow CH_3{-}CH_2{-}CH_2) + \mathrm{LOC}(CH_3{-}CH_2{-}CH_2) - \mathrm{LOC}(CH_3) - \mathrm{LOC}(CH_2{=}CH_2) \qquad (5)$$

where the LOC(x) terms on the right-hand side of this equation are those given in the last row of Table 2.

Similar formulas can be derived for atomization energies, ionization potentials, electron affinities, and barrier heights, although the last of these involves treating a rather more complex situation. This most recent work[5e] forms the basis for the method described here. Specifically, the accuracy of B3LYP's barrier height prediction was improved with simple numerical corrections to the reactant, product, and transition state energies.[5e] The success of this effort suggests that we can develop a robust description of a potential energy surface by interpolating these corrections between the various stationary points (reactant, transition state, and product) to arrive at corrections for points intermediate between stationary points. This is described in detail in the section that follows.

III. Development of Continuous Localized Orbital Corrections

III.A. An Overview of CLOCs. In this section and throughout the rest of the text, various new terms will be introduced. Therefore, we have defined these terms for convenience in addition to others that will be defined later, in Table 3. In order to develop corrections for the entire B3LYP reaction coordinate profile, it is necessary to first evaluate the accuracy of B3LYP with respect to high-level post-HF benchmarks along the entirety of the reaction coordinate. While a fairly large amount of benchmark data exists for thermochemical properties such as enthalpies of reaction and barrier heights in the literature (wherein only stationary points are required), there is a relative paucity of

Scheme 1. Reaction between Methyl Radical and Ethene to Give Propyl Radical

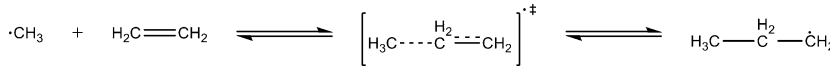


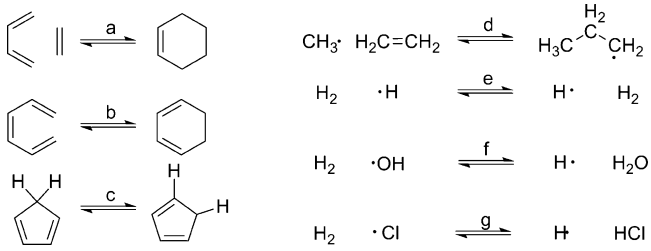
Table 2. Valence Bond Characteristics and Corresponding LOCs for the Reaction of Scheme 1

valence bond characteristic	LOC(x)	C_k[kcal/mol]	N_k methyl	ethene	propyl	transition state
sp^2 carbons	n/a	n/a	1	2	1	
$sp^{2.5}$ carbons	n/a	n/a				2
sp^3 carbons	n/a	n/a			2	1
C—H	$LOC(x)_{NPOLH}$	0.25	3	4	7	7
C—H attached to the radical-containing C	$LOC(x)_{RH}$	0.54	3		2	5
C---C	$LOC(x)_{MSBC/LSBC\_0.5}$	−2.05				1
C—C	$LOC(x)_{MSBC}$	−1.90			2	
C⋯C	$LOC(x)_{AA\_1.5}$	−0.88				1
C═C	$LOC(x)_{DBC}$	−1.00		1		
C—C adjacent to another C—C	$LOC(x)_{ESBC}$	−0.50			2	1/2
total LOC(x) [kcal/mol]			2.37	0.00	−1.97	1.27

Table 3. New Terms and Definitions

term	definition
active atom	an atom that belongs to at least one active bond
active bond	a bond with an order that changes throughout the reaction coordinate, for example, from a singe bond in the reactant to a double bond in the product
cutoff	the distance at which a bond is considered to have zero bond order according to the distances given in Table 4
inactive atom	an atom that belongs to no active bonds
inactive bond	a bond with an order which does not change throughout the reaction coordinate
product-side bond	description given to any bond which is intermediate in length between the transition state and product bonds
product-side structure	description given to any structure where the majority of the bonds are assigned as "product-side bonds"
reactant-side bond	description given to any bond which is intermediate in length between the transition state and reactant lengths
reactant-side structure	description given to any structure where the majority of the bonds are assigned as "reactant-side bonds"

Scheme 2. Reactions Employed in This Study[a]



[a] (a) Diels−Alder, (b) electrocyclic, (c) sigmatropic shift, (d) carbon radical, and (e−g) hydrogen transfer.

published data on complete reaction coordinates for systems larger than a few atoms. To address this problem, we produced our own curves computed with coupled cluster with single, double, and iterative triple excitations [CCSD(T)] for reactions a−f in Scheme 2. Specifically, single-point calculations along the reaction path were performed at the RCCSD-(T)/cc-pVTZ//B3LYP/6-31+G** level. CCSD(T) energies and B3LYP geometries were obtained with MolPro 2006.1[8] and Jaguar 7.6,[9] respectively. The data for reaction g were available in the literature.[10] The reactions were chosen from our latest DFT-LOC publication[5e] and represent a broad range of chemistries: (a) cycloaddition, (b) electrocyclization, (c) sigmatropic shift, (d) carbon radical, and (e−g) hydrogen radical reactions. Although hardly exhaustive, we argue that this set represents an acceptable starting point sufficient to evaluate the accuracy of our method as it applies to systems of general interest.

A subsequent examination of the B3LYP and post-HF profiles for each reaction gives a qualitative picture of what functional form the CLOC corrections should take. Specifically, the ideal CLOC is one that minimizes the error along the B3LYP-CLOC reaction profile in comparison with the post-HF profile for each arbitrary point x and hence is given by the following equation:

$$E_{\text{post-HF}}(x) = E_{\text{B3LYP-CLOC}}(x) = E_{\text{B3LYP}}(x) + \text{CLOC}(x) \tag{6}$$

When x is a stationary point, LOC(x) necessarily equals CLOC(x). Note that this equality only holds for stationary points, as LOC(x) is undefined for nonstationary point structures.

$$E_{\text{B3LYP-CLOC}}(x) = E_{\text{B3LYP-LOC}}(x) = E_{\text{B3LYP}}(x) + \text{CLOC}(x) = E_{\text{B3LYP}}(x) + \text{LOC}(x) \tag{7}$$

if and only if x is a stationary point. However, everywhere where x is not a stationary point, one must define CLOC(x).

A simple examination of the extant LOCs shows that they may be divided into those aimed at treating bonds, hybridization, radicals, hypervalency, environment, and charge transfer of any system x.

$$\text{LOC}(x) = \text{LOC}(x)_{\text{bond}} + \text{LOC}(x)_{\text{hyb}} + \text{LOC}(x)_{\text{radical}} + \text{LOC}(x)_{\text{hyperval}} + \text{LOC}(x)_{\text{environ}} + \text{LOC}(x)_{\text{CT}} \tag{8}$$

The continuous implementation necessarily takes the same form:

$$\text{CLOC}(x) = \text{CLOC}(x)_{\text{bond}} + \text{CLOC}(x)_{\text{hyb}} + \text{CLOC}(x)_{\text{radical}} + \text{CLOC}(x)_{\text{hyperval}} + \text{CLOC}(x)_{\text{environ}} + \text{CLOC}(x)_{\text{CT}} \tag{9}$$

Each term of eq 9 will be discussed in its own subsection directly following this one.

To begin the calculation of CLOC(x) for an arbitrary molecule x, we first require the availability of all relevant stationary points (reactant, transition state, and product) for reference. Specifically, the Cartesian coordinates of all of these structures obtained with the same level of theory as the arbitrary point (here, B3LYP/6-31+G**) must be provided.

In order to perform the interpolation between the stationary points, we must know where along the reaction coordinate profile the arbitrary structure lies with respect to the input structures. To this end, the arbitrary structure is analyzed against these input structures to determine whether it is reactant-side or product-side (i.e., whether it lies along the reaction coordinate connecting reactant to transition state, or transition state to product, respectively). Because of this, the quality of the user-provided stationary points is critical. Each bond of the structure x is analyzed individually with respect to its bond length l_x and receives its own assignment: either reactant- or product-side. A reactant-side bond is intermediate in length between the reactant and transition state lengths, i.e., $l_r \leq l_x < l_{ts}$ or $l_r \geq l_x > l_{ts}$. Similarly, a product-side bond is intermediate in length between the product and transition state lengths, i.e., $l_p \leq l_x < l_{ts}$ or $l_p \geq l_x > l_{ts}$. A structure that lies strictly along the reaction coordinate will have all bonds fall into the same category; however, this is not necessary for our algorithm to function, as each bond is interpolated independently. In spite of the ability to treat points that do not lie strictly along the reaction coordinate, in this work we have restricted ourselves to the study of structures that lie along the reaction coordinate exclusively. While we have high confidence in the ability of our method to treat these points, treatment of points that do not lie exactly along the reaction coordinate is feasible where these points lie at least close to the reaction coordinate. Because the integrity of the method outlined is dependent upon the choice of reaction coordinate, meaningful results may not be obtained for cases where the choice of most appropriate reaction coordinate is not straightforward. However, we leave an assessment of the accuracy of the model for such structures to a future publication.

Once a bond in x is determined to be either product-side or reactant-side, we use its bond length, l_x, to determine quantitatively where along that half of the reaction coordinate it lies. Each l_x is compared to the nearest equilibrium bond lengths, l_{eq} (reactant if it is a reactant-side bond, product if it is a product-side bond), and the transition-state bond length, l_{ts}, to obtain δ_x.

$$\delta_x = \frac{l_x - l_{eq}}{l_{ts} - l_{eq}} \tag{10}$$

From eq 10, δ_x approaches zero for bond lengths close to the equilibrium lengths (reactant or product) and approaches one for bond lengths similar to the transition state lengths. Equation 10 is undefined, however, where l_{eq} is infinitely long, i.e., when the bond is completely broken, in either the reactant or product structure. This difficulty is encountered in all intermolecular reactions. To circumvent this problem, we have defined an effective cutoff length, such that any bond with a length exceeding the cutoff is instead assigned the cutoff value. At this length, the bond is assigned a bond order of zero, and hence no bond corrections, CLOC(x)$_{bond}$ or LOC(x)$_{bond}$, are assigned to it, as bond corrections are only assigned for bonds with nonzero bond orders. For the reactions depicted in Scheme 2, we have arrived empirically at the cutoff lengths given in Table 4. However, an inspection of Scheme 2 shows that only a limited number of bond types are studied: H−X, where X = H, O, Cl, and C, and C−C. Therefore, we are forced to define cutoff lengths for bonds heretofore not studied. To do so, we note that the cutoff lengths given in Table 4 for any bond i correspond to roughly twice the transition state bond length for that same bond, $2l_{ts}$. Specifically, the average transition state bond length, l_{ts}^{avg}, for all H−X bonds (X = H, O, Cl, C) studied herein, is l_{ts}^{avg} = 1.2 Å, and the empirically determined cutoff of 2.2 Å = $1.9l_{ts}^{avg}$. Likewise, the average transition state bond length for all C−C bonds studied herein is l_{ts}^{avg} = 2.3 Å, and therefore the cutoff of 4.0 Å = $1.7l_{ts}^{avg}$. Therefore, all cutoff lengths for systems heretofore not studied are taken as 1.8 times the length of the bond in the transition state, $1.8l_{ts}$, assuming transferability of the empirically determined cutoff length trend. We are not barring the possibility of refinement of these cutoff values when more reaction profiles are explored in the future.

Table 4. Cutoff Lengths for l_{eq}

bond type	cutoff length (Å)
H−X (X = H, O, Cl, C)	2.2
C−C	4.0
all other bonds	$1.8l_{ts}$

Equipped with our estimate of how far along the reaction coordinate the arbitrary structure lies with respect to each bond length (δ_x), we only need the equilibrium LOCs, to proceed with the interpolations between LOC(reactant), LOC(ts), and LOC(product). These are determined using an automated script that gives the corrections already described in previous publications[5c,e] with one minor exception described in section III.F.

The obtained equilibrium LOCs are then used to calculate each component i of CLOC(x) (see eq 9) according to the equation

$$\mathrm{CLOC}(\delta_x)_i = \mathrm{LOC(eq)}_i + f(\delta_x)_i \times \Delta\mathrm{LOC}_i \tag{11}$$

where

$$\Delta\mathrm{LOC}_i = \mathrm{LOC(ts)}_i - \mathrm{LOC(eq)}_i \tag{12}$$

and LOC(eq)$_i$ is set to be LOC(reactant)$_i$ for a reactant-side interpolation or LOC(product)$_i$ for a product-side interpolation. Therefore, we are only left with the task of choosing the appropriate $f(\delta_x)_i$ for each component i of eq 9, where i can be bond, hyb, etc.

In eq 11, the i subscript is used to emphasize that we have chosen to interpolate each CLOC term individually, each term receiving its own unique $f(\delta_x)_i$. Hypothetically, the interpolations could be performed instead on the basis of just one value of $f(\delta_x)$ that reflects where along the reaction coordinate the structure lies in its entirely. However, in spite

of its simplistic appeal, to obtain meaningful results with this method, all bond lengths and hybridization states etc. must fall at the same place along the reaction coordinate. By using the formulation presented in eq 11, where each term is interpolated individually, no such restriction is applied. Therefore, we have chosen to interpolate on a term-by-term basis to allow for increased flexibility and accuracy.

Clearly, the ability to assign LOCs to stationary points, and hence interpolate CLOCs for all intermediate structures, is dependent upon the ability to accurately assign Lewis structures to the former. All assignments of Lewis structures in this work were performed using an automated script (also used in other works described above[5c,e]). When this automatic assignment fails, human intervention might be necessary to provide information about the formal charges and/or spins in the same input file with the input structures, thereby preventing misassignment of more complicated systems. It is possible that the CLOC approach will be inapplicable to some systems with a poorly understood or badly defined Lewis structure. For the vast majority of systems of practical interest, no difficulty is encountered in this respect whatsoever. Specifically, large systems, such as those of interest to organic chemists and biochemists, are regularly studied using DFT for its excellent balance of performance and accuracy.[11] These same systems generally have well-defined Lewis structures and can therefore be treated easily with our CLOC methodology, as shown by the successful treatment of various organic chemistry reactions in our latest work.[5e]

III.B. CLOCs for Bonds, CLOC$(x)_{bond}$. As is discussed in detail in our previous publications,[5] the DFT-LOC model provides improvements to the estimation of nondynamical electron correlation by a specific DFT functional for localized electron pairs. The DFT-LOC bond corrections, or LOC$(x)_{bond}$, rest upon the assumption that the localized nuclear framework supporting an electron pair is a principal factor controlling the deviations in value of the nondynamical correlation from the "average" value within global hybrid functionals. Therefore, empirical corrections are applied on the basis of these localized frameworks. Consider, for example, the corrections applied to single bonds between heavy atoms of various lengths, when the 6-311++G(3df,3pd) basis set is used: short (−1.36 kcal/mol), medium (−1.90 kcal/mol), and long (−2.57 kcal/mol). These values become appreciably more negative with increasing bond length. This reflects the physically intuitive notion that as bond length increases, nondynamical correlation becomes more negative (as the electrons have more room to avoid each other), and B3LYP systematically underestimates this two-particle correlation effect with increasing severity.

The total LOC$(x)_{bonds}$ is given as the sum of all the various terms' bond LOCs:

$$\mathrm{LOC}(x)_{\mathrm{bonds}} = \sum_i \mathrm{LOC}(x)_i \tag{13}$$

where i runs over the 14 LOCs unique to bonds. The rationale for each correction is described in detail in ref 5, while the optimized value for each correction is given in the Supporting Information.

Analogously, the total CLOC$(x)_{bonds}$ is given by

$$\mathrm{CLOC}(x)_{\mathrm{bonds}} = \sum_i \mathrm{CLOC}(x)_i \tag{14}$$

where i again runs over the 14 LOCs unique to bonds.

All LOCs for bonds, LOC$(x)_{bonds}$, are designed to treat bonds of order 0.5, 1, 1.5, 2, 2.5, and 3. Yet we desire the ability to treat all bond orders and, thereby, transform LOC$(x)_{bond}$ to CLOC$(\delta_x)_{bond}$. As stated previously, CLOC$(\delta_x)_i$ is given by eq 11, which is modified such that it is specific to CLOC$(\delta_x)_{bond}$ (i = bond) and is written as

$$\mathrm{CLOC}(\delta_x)_{\mathrm{bond}} = \mathrm{LOC(eq)}_{\mathrm{bond}} + f(\delta_x)_{\mathrm{bond}} \times \Delta\mathrm{LOC}_{\mathrm{bond}} \tag{15}$$

where

$$\Delta\mathrm{LOC}_{\mathrm{bond}} = \mathrm{LOC(ts)}_{\mathrm{bond}} - \mathrm{LOC(eq)}_{\mathrm{bond}} \tag{16}$$

Therefore, we are only left with the task of choosing a proper form for $f(\delta_x)_{bond}$ such that it satisfies the appropriate boundary conditions:

$$f(\delta_x)_{\mathrm{bond}} = \begin{cases} 0, & \text{if } \delta_x = 0; \\ 1, & \text{if } \delta_x = 1 \end{cases} \tag{17}$$

Inspection shows that these boundary conditions are designed to ensure that CLOC$(\delta_x)_{bond}$ = LOC(eq)$_{bond}$ at δ_x = 0 (i.e., at the reactant or product) and that CLOC$(\delta_x)_{bond}$ = LOC(ts)$_{bond}$ at δ_x = 1 (i.e., at the transition state). Put simply, we are ensuring agreement between the previously developed LOCs and the continuous version, CLOCs, in the reactant, product, and transition state; i.e., LOC(x) = CLOC(x) where x is a stationary state.

It is reasonable to define $f(\delta_x)_{bond}$ as either a linear, Gaussian, or power function to satisfy these boundary conditions:

$$f(\delta_x)_{\mathrm{bond}} = \delta_x \tag{18}$$

$$f(\delta_x)_{\mathrm{bond}} = \mathrm{e}^{-\gamma(1-\delta_x)^2} \tag{19}$$

$$f(\delta_x)_{\mathrm{bond}} = 1 - (1 - \delta_x)^n \tag{20}$$

One can envision using other functions as well to perform the interpolations. For example, $f(\delta_x) = \sin(\delta_x \times \pi/2)$ could also be employed. We are not barring the possibility of adopting this or other interpolating functions in the future.

While eqs 18−20 all satisfy the necessary boundary conditions, it is also necessary that any $f(\delta_x)$ be everywhere differentiable such that its gradients can be defined. It is easy to see how linear interpolations based upon eq 18 would lead to nondifferentiable cusps at the transition state (δ_x = 1), where the reactant-side and product-side linear interpolations intersect, giving a curve with a shape similar to a triangle wave. For this reason, linear interpolations were discarded in spite of their simplicity. Among the power functions described by eq 20, we found cubic functions to best mimic the qualitative shape of the B3LYP error (for at least the shorter-bond-length half of the reaction coordinate) and therefore to give the best results. Unfortunately, testing has revealed that for the application of the CLOC method, cubic functions do not decay quickly enough to zero as δ_x

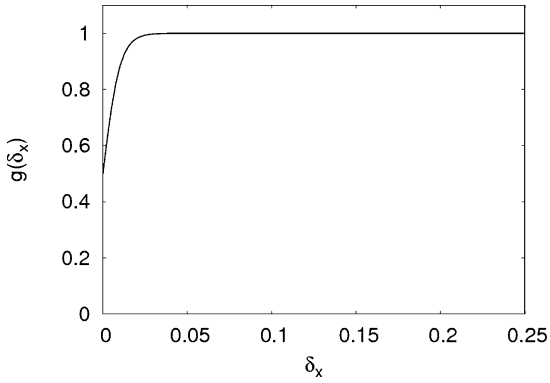


Figure 3. $g(\delta_x)$ vs δ_x as defined by eq 21.

approaches zero. To solve this problem, we multiply our cubic function by a function of the following form:

$$g(\delta_x) = \frac{1}{1 + e^{-\beta\delta_x}} \tag{21}$$

where β is an adjustable coefficient which controls the rate of decay, here chosen to be 200, and δ_x is defined by eq 10.

Inspection of Figure 3 shows that $g(\delta_x)$ decays rapidly as $\delta_x \rightarrow 0$. Therefore, multiplication of the cubic function given in eq 20 with $g(\delta_x)$ gives a new function that decays to zero with the proper rate as $\delta_x \rightarrow 0$. This function now has a desirable analytical behavior and can be used to interpolate between our limiting stationary point LOCs according to eq 11 for all bonds. Further, Gaussian functions, given by eq 19, can also be used, without modification, to the same end.

Interestingly, we have found that a combination of the two functions, modified cubic and Gaussian, serves as an even better match for the DFT B3LYP error as a function of intrinsic reaction coordinate. To understand how the two functions are combined, first consider how an active bond changes along the reaction coordinate. The characteristics of a bond in a transition state structure along the reaction coordinate change from bond order s_r with length l_r in the reactant to bond order s_p with length l_p in the product. For such a bond, we assume that the bond order in the transition state is an average of these two bond orders, $s_{ts} = (s_r + s_p)/2$, with corresponding length l_{ts}.

If $l_r > l_p$, any bond in the arbitrary structure with bond order s_x and bond length l_x has an interpolated LOC obtained from eq 11 where $f(\delta_x)_{bond}$ is given by

$$f(\delta_x)_{bond} = \begin{cases} e^{-\gamma(1-\delta_x)^2}, & \text{if } l_r > l_x \geq l_{ts}; \\ \dfrac{1-(1-\delta_x)^3}{1+e^{-\beta\delta_x}}, & \text{if } l_{ts} > l_x \geq l_p \end{cases} \tag{22}$$

Alternatively, if $l_r < l_p$, we have

$$f(\delta_x)_{bond} = \begin{cases} \dfrac{1-(1-\delta_x)^3}{1+e^{-\beta\delta_x}}, & \text{if } l_r > l_x \geq l_{ts}; \\ e^{-\gamma(1-\delta_x)^2}, & \text{if } l_{ts} > l_x \geq l_p \end{cases} \tag{23}$$

In both eqs 22 and 23, γ is an adjustable parameter to modulate the width of the Gaussian curve, set here to 5. β is used as in eq 21, and δ_x is defined by eq 10.

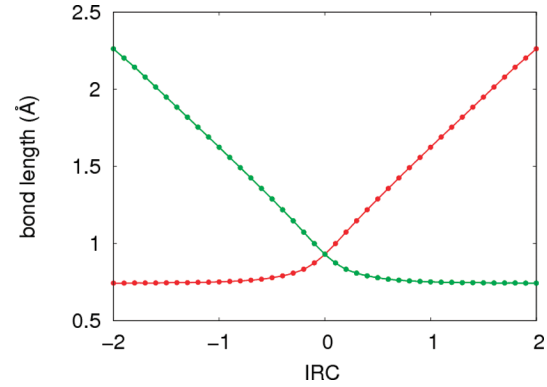


Figure 4. Bond lengths vs intrinsic reaction coordinate (IRC) for $H^1{-}H^2 + H^3 \rightarrow H^1 + H^2{-}H^3$. $H^1{-}H^2$ bond lengths are shown in red, while $H^2{-}H^3$ bond lengths are shown in green. Note that $H^1{-}H^2$ bond length changes rapidly for IRC > 0, yet slowly for IRC < 0. The opposite is true for $H^2{-}H^3$ according to symmetry.

Although empirically derived, this differential treatment, i.e., cubic interpolations for the shorter-bond-length half of the reaction coordinate (i.e., $l_r < l_x < l_{ts}$ or $l_{ts} > l_x > l_p$) and Gaussian interpolations for the longer, is founded upon the dependence of the reaction coordinate on bond length. Consider the reaction $H^1{-}H^2 + H^3 \rightarrow H^1 + H^2{-}H^3$, where each hydrogen has been marked with a unique superscript for the purpose of the argument. Inspection of Figure 4 shows that the $H^1{-}H^2$ bond length changes insignificantly with reaction coordinate on the reactant side, from 0.74 Å in the reactant to 0.93 Å in the transition state. Conversely, the $H^1{-}H^2$ bond length changes considerably with reaction coordinate on the product side, from 0.93 Å in the transition state to essentially infinite bond length in the product. Therefore, a Gaussian is employed on the product side for $H^1{-}H^2$ interpolations to ensure that the CLOC decays rapidly along the reaction coordinate, whereas a cubic function is used on the reactant side for $H^1{-}H^2$ interpolations for the opposite reason.

III.C. Defining CLOCs for Hybridization, CLOC$(x)_{hyb}$. In previous publications, we have also defined LOCs to describe various hybridization states, LOC$(x)_{hyb}$. While it is true that DFT in general benefits from significant cancellation of intra-atomic error as one goes from atoms to a molecule, i.e., as bonds are formed and atomic electronic structure is changed, these parameters were developed to address errors that remain in spite of this cancellation. These LOC$(x)_{hyb}$ parameters address the relatively large changes in orbital sizes, shapes, and occupancies that accompany bond formation and hence cause variations in nondynamical electron correlation for an electron pair contained in an orbital with a particular hybridization.

Each of these LOCs has its own unique purpose. For example, LOC$(x)_{N/P_sp2}$ and LOC$(x)_{N/P_sp3}$ are assigned for each nitrogen or phosphorus atom with sp^2 or sp^3 hybridization, respectively. Extensive definitions and optimized values (C_k) for all hybridization LOCs [LOC$(x)_{hyb}$] can be found in our previous publications[5] or the Supporting Information.

If we wish to treat hybridization states other than sp, $sp^{1.5}$, sp^2, $sp^{2.5}$, and sp^3, we must transform our LOC$(x)_{hyb}$ to the continuous CLOC$(x)_{hyb}$. To simplify the calculation of

$\mathrm{CLOC}(x)_{\mathrm{hyb}}$, we have split each of the $\mathrm{CLOC}(x)_{\mathrm{hyb}}$ terms into "active" and "inactive" terms so that $\mathrm{CLOC}(x)_{\mathrm{hyb}}$ is written as the sum of these two components.

$$\mathrm{CLOC}(x)_{\mathrm{hyb}} = \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}} + \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{inactive}} \quad (24)$$

The total $\mathrm{CLOC}(x)_{\mathrm{hyb}}$ is given by the sum of all hybridization CLOCs, each having an inactive and active component:

$$\mathrm{CLOC}(x)_{\mathrm{hyb}} = \sum_i [\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}} + \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{inactive}}]_i \quad (25)$$

where i runs over the nine LOCs unique to hybridization.

Accordingly, a given reaction profile (with reactant, transition state, and product structures) is processed to classify all of the bonds as either "active" or "inactive", i.e., as bonds with changed or unchanged bond orders along the reaction coordinate, respectively. Similarly, atoms are also classified as either inactive or active. Inactive atoms are those attached exclusively to inactive bonds, whereas active atoms are those attached to one or more active bonds. For example, in the reaction between methanol and the hydrogen atom, $H_3C-O-H + H\cdot \rightarrow H_3C-O\cdot + H-H$, the OH bond and the HH bond are both active, whereas the CO and CH bonds are inactive. Further, the carbon atom and hydrogens attached to it are inactive atoms, while all others are active.

The inactive component of $\mathrm{CLOC}(x)_{\mathrm{hyb}}$ is taken in direct analogy to eq 1 as

$$\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{inactive}} = \sum_k N_k C_k \quad (26)$$

where the index k runs over all inactive atoms, assigning LOCs in the same way as if these atoms were part of an equilibrium structure. Because inactive atoms experience no change in hybridization throughout the entirety of the reaction coordinate (as ascertained upon analysis of the reactant, product, and transition state structures input), we treat them as if they were still in their equilibrium states. Instead, we concern ourselves with treating only the active components of the reaction coordinate for hybridization in a dynamic fashion, $\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$. Specifically, $\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$ is taken as an interpolated value between $\mathrm{LOC(eq)}_{\mathrm{hyb}}$ and $\mathrm{LOC(ts)}_{\mathrm{hyb}}$, where $\mathrm{LOC(eq)}_{\mathrm{hyb}}$ is $\mathrm{LOC(r)}_{\mathrm{hyb}}$ for a reactant-side interpolation or $\mathrm{LOC(p)}_{\mathrm{hyb}}$ for a product-side interpolation. Adapting eq 11 to the active hybridization term gives

$$\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}} = \mathrm{LOC(eq)}_{\mathrm{hyb}}^{\mathrm{active}} + f(\delta_x)_{\mathrm{hyb}} \times \Delta\mathrm{LOC}_{\mathrm{hyb}}^{\mathrm{active}} \quad (27)$$

where

$$\Delta\mathrm{LOC}_{\mathrm{hyb}}^{\mathrm{active}} = \mathrm{LOC(ts)}_{\mathrm{hyb}}^{\mathrm{active}} - \mathrm{LOC(eq)}_{\mathrm{hyb}}^{\mathrm{active}} \quad (28)$$

In order to properly specify $\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$, we must first arrive at a proper definition of hybridization itself. While a bond is defined simply by two atomic centers and the distance between them, l, hybridization of an atom is a more complex characteristic which depends on all atoms surrounding the given atom, as well as the respective bond lengths, l_1, l_2, ..., l_n. Therefore, eqs 22 and 23, which depend only upon one bond length, are not sufficient to define hybridization, and similarly to the situation above, we define an interpolating $f(\delta_x)_{\mathrm{hyb}}$ for eq 27 such that it satisfies the appropriate boundary conditions.

$$f(\delta_x)_{\mathrm{hyb}} = \begin{cases} 0, & \text{for stationary state;} \\ 1, & \text{for transition state} \end{cases} \quad (29)$$

As in section III.B for $\mathrm{CLOC}(x)_{\mathrm{bonds}}$, a simple examination reveals that these boundary conditions ensure that $\mathrm{LOC(ts)}_{\mathrm{hyb}}^{\mathrm{active}} = \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$ where $f(\delta_x)_{\mathrm{hyb}} = 1$, i.e., at the transition state, and that $\mathrm{LOC(eq)}_{\mathrm{hyb}}^{\mathrm{active}} = \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$ where $f(\delta_x)_{\mathrm{hyb}} = 0$, i.e., for the reactant or product. Again, we are simply ensuring that $\mathrm{LOC}(x)_{\mathrm{hyb}}^{\mathrm{active}} = \mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$ when x is a stationary state.

To take into account the multiatom dependence of hybridization, we have defined $f(\delta_x)_{\mathrm{hyb}}$ for atomic LOCs as the average of all $f(\delta_x)_{\mathrm{bond}}$'s values over all active bonds connected to that atomic center:

$$f(\delta_x)_{\mathrm{hyb}} = \frac{1}{n}\sum_{i=1}^{n} [f(\delta_x)_{\mathrm{bond}}]_i \quad (30)$$

where i is an index that runs over all active bonds. It is easy to see that as the bonds connected to any particular atom become more transition-state-like, as $f(\delta_x)_{\mathrm{bond}} \rightarrow 1$, on average, the interpolated hybridization also becomes more transition-state-like, that is, $f(\delta_x)_{\mathrm{hyb}} \rightarrow 1$. This also holds in the reverse direction, i.e., as bonds become more reactant- or product-like. In this manner, the interpolated value of $\mathrm{CLOC}(x)_{\mathrm{hyb}}^{\mathrm{active}}$ according to eq 27 is tuned to reflect how reactant-, product-, or transition-state-like the hybridization of an active atom is as a function of how reactant-, product-, or transition-state-like the bonds connected to it are on average.

III.D. Defining CLOCs for Radicals, $\mathrm{CLOC}(x)_{\mathrm{radical}}$. As argued extensively in our previous publications,[5] the self-interaction term in DFT is used to quantitatively model the nondynamical electron correlation of an electron pair. However, this self-interaction term becomes problematic for unpaired electrons. We have previously developed corrections to specifically treat atoms with radicals localized on them, $\mathrm{LOC}(x)_{\mathrm{radical}}$, to remedy systematic overbinding: $\mathrm{LOC}(x)_{\mathrm{RH}}$, $\mathrm{LOC}(x)_{\mathrm{RA}}$, and $\mathrm{LOC}(x)_{\mathrm{RT}}$, to treat atomic centers with localized radicals that have neighboring bonds to hydrogen, single or double bonds to heavy atoms, or triple bonds to heavy atoms, respectively.

The number of unpaired electrons and formal charge on each atomic center are ascertained by assuming a set number of valence electrons, χ, for each atom type, as specified in Table 5. The formal charge (q) and number of unpaired electrons (u) is then a function of the element's group number

Table 5. Valency for Each Atom Type

atom type	χ	example
H, He	2	H_2
Al, B	6	BH_3
Cl, P, or S with $\sum_i s_i > 8 - g$	10	PCl_5
all other 1st and 2nd row atoms	8	CH_4

on the periodic table (g) and the bond order of all bonds connected to it (s_i) according to

$$g - q + u + \sum_{i=1}^{n} s_i = \chi \tag{31}$$

Given the number of unpaired electrons, u, we can compute the radical CLOCs: CLOC$(x)_{RH}$, CLOC$(x)_{RA}$, and CLOC$(x)_{RT}$. [For a complete list of all CLOCs and their definitions and values, including the radical CLOCs, CLOC$(x)_{radical}$, the reader is referred to our previous publications[5] or the Supporting Information.]

According to eq 1, the contribution due to radical LOCs to the total LOC for an equilibrium structure will be given by

$$\mathrm{LOC}(x)_{\mathrm{radical}} = \sum_i \mathrm{LOC}(x)_i = \sum_i N_i \times C_i \tag{32}$$

where, here, i runs over the three LOCs unique to radicals [LOC$(x)_{RH}$, LOC$(x)_{RA}$, and LOC$(x)_{RT}$] and N_k and C_k are the count and LOC value, respectively. Specifically,

$$N_{\mathrm{RH}} = \sum_i u_i[\eta_{\mathrm{RH}}]_i \tag{33}$$

$$N_{\mathrm{RA}} = \sum_i u_i[\eta_{\mathrm{RA}}]_i \tag{34}$$

$$N_{RT} = \sum_i u_i[\eta_{\mathrm{RT}}]_i \tag{35}$$

where u_i is the number of unpaired electrons, $[\eta_{RH}]_i$ is the number of single bonds to hydrogen atoms, $[\eta_{RA}]_i$ is the number of single or double bonds to non-hydrogen atoms, and $[\eta_{RT}]_i$ is the number of triple bonds, all corresponding to center i.

In order to adapt our equilibrium LOC$(x)_{radical}$ contribution to the continuous representation, CLOC$(x)_{radical}$, we must allow for noninteger values of n_{RH}, n_{RA}, and n_{RT}. To accomplish this, we substitute η in eqs 33−35 with some function of the bond orders for the bonds of interest, $f(s)$:

$$N_{\mathrm{RH}} = \sum_i u_i[f_{\mathrm{RH}}(s_{\mathrm{RH}})]_i \tag{36}$$

$$N_{\mathrm{RA}} = \sum_i u_i[f_{\mathrm{RA}}(s_{\mathrm{RA}})]_i \tag{37}$$

$$N_{\mathrm{RT}} = \sum_i u_i[f_{\mathrm{RT}}(s_{\mathrm{RT}})]_i \tag{38}$$

where s_{RH}, s_{RA}, and s_{RT} are the bond orders between the atomic center of interest i and its neighboring hydrogen atom (RH) or neighboring non-hydrogen atom (RA or RT).

Again, we desire $f(s)$ functions that both are differentiable and satisfy the appropriate boundary conditions. LOC$(x)_{RH}$ is applied to any radical-containing atom i with bonds to hydrogen. Therefore, the boundary conditions dictate that there be no LOC$(x)_{RH}$ when atom center i is not bonded to a hydrogen atom. Conversely, there must be one LOC$(x)_{RH}$ applied for each (single) bond to hydrogen from atom center i. The boundary condition for $f_{RH}(s_{RH})$ in eq 36 is hence given by

$$f_{\mathrm{RH}}(s_{\mathrm{RH}}) = \begin{cases} 0, & \text{for } s_{\mathrm{RH}} = 0 \text{ or } s_{\mathrm{RH}} = 2; \\ 1, & \text{for } s_{\mathrm{RH}} = 1 \end{cases} \tag{39}$$

where s_{RH} is the bond order of the bond between atom center i and the neighboring hydrogen atom.

Likewise, LOC$(x)_{RA}$ is applied to any radical-containing atom i with single or double bonds to non-hydrogen atoms. Thus, we desire one LOC$(x)_{RA}$ for each atom center i with a single or double bond to another non-hydrogen atom and no LOC$(x)_{RA}$ for each atom center i with no bond or a triple bond to another non-hydrogen atom. The boundary condition for $f_{RA}(s_{RA})$ in eq 37 is hence given by

$$f_{\mathrm{RA}}(s_{\mathrm{RA}}) = \begin{cases} 0, & \text{for } s_{\mathrm{RA}} = 0 \text{ or } s_{\mathrm{RA}} = 3; \\ 1, & \text{for } s_{\mathrm{RA}} = 1 \text{ or } s_{\mathrm{RA}} = 2 \end{cases} \tag{40}$$

where s_{RA} is the bond order between the radical-containing atom i and the neighboring non-hydrogen atom.

Lastly, LOC$(x)_{RT}$ is applied to any radical-containing atom i with a triple bond to a non-hydrogen atom. Thus, we desire one LOC$(x)_{RT}$ for each atom center i with a triple bond to another non-hydrogen atom and no LOC$(x)_{RT}$ for each atom center i with no triple bond to another non-hydrogen atom. The boundary condition for $f_{RT}(s_{RT})$ in eq 38 is hence given by

$$f_{\mathrm{RT}}(s_{\mathrm{RT}}) = \begin{cases} 0, & \text{for } s_{\mathrm{RT}} = 2 \text{ or } s_{\mathrm{RT}} = 4; \\ 1, & \text{for } s_{\mathrm{RT}} = 3 \end{cases} \tag{41}$$

where s_{RT} is the bond order between the radical-containing atom i and the neighboring non-hydrogen atom.

As discussed above, we can readily employ Gaussian functions to both meet the differentiability requirement and satisfy our various boundary conditions.

$$f_{\mathrm{RH}}(s_{\mathrm{RH}}) = \mathrm{e}^{-\upsilon(s_{\mathrm{RH}}-1)^2} \tag{42}$$

$$f_{\mathrm{RA}}(s_{\mathrm{RA}}) = \begin{cases} \mathrm{e}^{-\upsilon(s_{\mathrm{RA}}-1)^2}, & \text{if } s_{\mathrm{RA}} \le 1; \\ 1, & \text{if } 1 < s_{\mathrm{RA}} \le 2; \\ \mathrm{e}^{-\upsilon(s_{\mathrm{RA}}-2)^2}, & \text{if } s_{\mathrm{RA}} > 2 \end{cases} \tag{43}$$

$$f_{\mathrm{RT}}(s_{\mathrm{RT}}) = \mathrm{e}^{-\upsilon(s_{\mathrm{RT}}-3)^2} \tag{44}$$

where υ is set to 5 to ensure a proper rate of growth and decay for our functions.

Therefore, the CLOC$(x)_{radical}$ term is written as

$$\mathrm{CLOC}(x)_{\mathrm{radical}} = \mathrm{CLOC}(x)_{\mathrm{RH}} + \mathrm{CLOC}(x)_{\mathrm{RA}} + \mathrm{CLOC}(x)_{\mathrm{RT}} = N_{\mathrm{RH}} \times C_{\mathrm{RH}} + N_{\mathrm{RA}} \times C_{\mathrm{RA}} + N_{\mathrm{RT}} \times C_{\mathrm{RT}} \tag{45}$$

where N_{RH}, N_{RA}, and N_{RT} are defined by eqs 36−38 and 42−44.

III.E. CLOCs for Hypervalency, CLOC$(x)_{hyperval}$. We also define LOCs for atoms with more than eight valence electrons or two valence electrons, for hydrogen and helium atoms: LOC$(x)_{hyperval}$. These LOCs are designated LOC-

$(x)_{OCT\_EXP}$ and $LOC(x)_{H\_dival}$, respectively. The total $LOC(x)_{hyperval}$ is thus given by the sum of these two terms:

$$LOC(x)_{hyperval} = LOC(x)_{OCT\_EXP} + LOC(x)_{H\_dival} \quad (46)$$

Analogously, we write the continuous version, $CLOC(x)_{hyperval}$, as

$$CLOC(x)_{hyperval} = CLOC(x)_{OCT\_EXP} + CLOC(x)_{H\_dival} \quad (47)$$

Both of these terms will be discussed in the subsections that follow.

III.E.1. Hydrogen Hypervalency: CLOC(x)$_{H\_dival}$. In ref 5e, 105 transition states and barrier heights were analyzed at the B3LYP level. A thorough analysis of the errors in these B3LYP barrier heights reveals that transition states in which the central hydrogen atom is flanked by at least one nonhydrogen/noncarbon atom (as shown in the transition states of examples i−iii below) all display systematic errors. Presumably, this originates in overestimation of nondynamical electron correlation due to localized high electron density.

In this same study,[5e] we also found that the transition state in which the central hydrogen atom is flanked by two hydrogen atoms (as in example iv below) displays approximately the same error in barrier height. This transition state is highly analogous to the well-described H_2^+ molecule[12] where the self-interaction term described earlier does not serve to model nondynamical electron correlation but instead engenders a clear source of systematic error.[13] Accordingly, $LOC(x)_{H\_dival}$ is assigned to cases where the central hydrogen atom is flanked by two hydrogens to remedy this error as well.

In summary, $LOC(x)_{H\_dival}$ is applied to transition states in which the central hydrogen atom is flanked by at least one noncarbon/nonhydrogen atom ($n_A \geq 1$, in eq 50) or where the central hydrogen atom is flanked by two additional hydrogen atoms ($n_H = 2$ in eq 50).

We may rewrite the above discussion in terms of equations as follows. In the discrete version of the approach, the H_dival correction for any system x, $LOC(x)_{H\_dival}$, is written as

$$LOC(x)_{H\_dival} = N_{H\_dival} \times C_{H\_dival} \quad (48)$$

C_{H_dival} is the correction's numerical value, optimized to reduce the B3LYP error, and N_{H_dival} is given by

$$N_{H\_dival} = \sum_i \eta_i \quad (49)$$

where i is an index that runs over all hydrogen atoms and η is defined by the number of bonds to each hydrogen atom according to

$$\eta = \begin{cases} 0, & \text{if } n_A < 1; \\ 1, & \text{if } n_A \geq 1 \text{ or } n_H = 2 \end{cases} \quad (50)$$

Here, n_A is the number of bonds between hydrogen atom i and noncarbon/non-hydrogen atoms, whereas n_H is the number of bonds between hydrogen atom i and other hydrogens. For example, the transition states of the following reactions each merit $\eta = N_{H\_dival} = 1$:

(i) $H_2O + NH_2 \rightarrow HO + NH_3$ via $[HO{-}H{-}NH_2]^{\bullet\ddagger}$
(ii) $CH_4 + OH \rightarrow CH_3 + OH_2$ via $[CH_3{-}H{-}OH]^{\bullet\ddagger}$
(iii) $H_2 + Cl \rightarrow H + HCl$ via $[H{-}H{-}Cl]^{\bullet\ddagger}$
(iv) $H_2 + H \rightarrow H + H_2$ via $[H{-}H{-}H]^{\bullet\ddagger}$

In each of these transition states, the central hydrogen atom is flanked by either two hydrogen atoms or at least one nonhydrogen/noncarbon atom. Alternatively, the transition states of the following reactions have $\eta = N_{H\_dival} = 0$:

(v) $CH_4 + CH_3 \rightarrow CH_3 + CH_4$ via $[CH_3{-}H{-}CH_3]^{\bullet\ddagger}$
(vi) $CH_3 + H_2 \rightarrow CH_4 + H$ via $[CH_3{-}H{-}H]^{\bullet\ddagger}$

In these transition states, the central hydrogen atom is flanked by either two carbon atoms or one carbon atom and one hydrogen atom. Notice that the case where the central hydrogen atom is flanked by two hydrogen atoms (example iv) still merits $\eta = N_{H\_dival} = 1$ as described in the discussion above.

We assume integer bond orders in reactants and products and integer bond orders in addition to bond orders 0.5, 1.5, and 2.5 in transition states. In eq 50, any of these would be considered "bonds" and hence contribute to n as illustrated in the examples above.

To adopt $LOC(x)_{H\_dival}$ to the continuous case and hence specify the functional form for $CLOC(x)_{H\_dival}$, we redefine η as a function of δ_x. Specifically, eq 50 may be written as η_{X-H-Y}, where the subscripts X and Y indicate the type of atoms flanking the central hydrogen atom. As before, A is defined as any atom other than carbon or hydrogen.

$$\eta_{X-H-Y} = \begin{cases} \sqrt{\delta_{X-H} \times \delta_{Y-H}}, & \text{for } X = A_1 \text{ and } Y = A_2, \\ & \text{or } X = A \text{ and } Y = C, \\ & \text{or } X = A \text{ and } Y = H, \\ & \text{or } X = H_1 \text{ and } Y = H_2; \\ 0, & \text{for } X = C_1 \text{ and } Y = C_2, \\ & \text{or } X = C \text{ and } Y = H \end{cases} \quad (51)$$

In eq 51, δ_{X-H} is dependent upon the length between atom X and the central hydrogen atom, l_{X-H}, according to eq 10. δ_{Y-H} is defined similarly. Note that η_{X-H-Y} is defined for examples i−iv above but is always zero for examples v−vi. This is consistent with the prescriptions detailed at the beginning of this section.

An inspection of eq 51 shows that as both bond lengths, l_{X-H} and l_{Y-H}, approach their transition state lengths, δ_{X-H} and δ_{Y-H}, $\eta_{X-H-Y} \rightarrow 1$ and hence $CLOC(x)_{H\_dival}$ is applied. Likewise, as both bond lengths approach the equilibrium lengths, δ_{X-H} and δ_{Y-H}, $\eta_{X-H-Y} \rightarrow 0$ and hence $CLOC(x)_{H\_dival}$ is not applied. This is consistent with our understanding of hydrogen abstraction reactions, wherein the hydrogen being abstracted is divalent in the transition state, where $CLOC(x)_{H\_dival}$ is applied, but only monovalent in the reactant and product, where $CLOC(x)_{H\_dival}$ is not applied.

In eq 51, a product of δ_{X-H} and δ_{Y-H} is employed to ensure that η_{X-H-Y} is a function of both bond lengths l_{X-H} and l_{Y-H} and that it decays quickly to zero when at least one of the bond lengths is greater than the transition state bond length, i.e., as δ_{X-H} or $\delta_{Y-H} \rightarrow 0$. This reflects the fact that hypervalency on hydrogen is a function of two bond lengths. For example, if only l_{X-H} is near the transition state bond

length, giving $\delta_{X-H} \approx 1$, but l_{Y-H} is near the equilibrium length, giving $\delta_{Y-H} \approx 0$, in fact the central hydrogen is not hypervalent via chemical intuition. Accordingly, we desire $\eta_{X-H-Y} \approx 0$ such that $CLOC(x)_{H\_dival}$ is effectively not applied. This is indeed realized with the functional form of eq 51.

Alternatively, had we defined η_{X-H-Y} as simply the average of δ_{X-H} and δ_{Y-H}, i.e., $\eta_{X-H-Y} = (\delta_{X-H} + \delta_{Y-H})/2$, the proper behavior would not be observed in the illustrative example given above. Specifically, for $\delta_{X-H} \approx 1$ and $\delta_{Y-H} \approx 0$, $\eta_{X-H-Y} = (\delta_{X-H} + \delta_{Y-H})/2 \approx 1/2$ and $CLOC(x)_{H\_dival}$ would be nonzero and hence imply partial hypervalency on the central hydrogen atom, in spite of the fact that we know from chemical intuition that hypervalent character of this central hydrogen atom is negligible.

The necessity for the square root over the product in eq 51 becomes clear upon consideration of an additional illustrative example. Imagine a central hydrogen atom that is only "half-hypervalent", i.e., halfway along the reaction coordinate between the reactant or product and transition state in a standard hydrogen abstraction reaction. Here, $\delta_{X-H} = \delta_{Y-H} = 1/2$, and we desire $\eta_{X-H-Y} = 1/2$ to reflect this "half-hypervalency." Defining η_{X-H-Y} as a simple product, i.e., $\eta_{X-H-Y} = \delta_{X-H} \times \delta_{Y-H}$, would yield $\eta_{X-H-Y} = 1/4$ in this case, and hence this system would not be described as "half-hypervalent" as we desire but instead as only "quarter-hypervalent". Instead, we define η_{X-H-Y} as the square of the product to effect the proper behavior, in this case, $\eta_{X-H-Y} = 1/2$, consistent with our understanding that this system is "half-hypervalent".

Combining eq 49 with eq 51 allows us to define $CLOC(x)_{H\_dival}$ as

$$CLOC(x)_{H\_dival} = N_{H\_dival} \times C_{H\_dival} \tag{52}$$

III.E.2. Heavy-Atom Hypervalency: $CLOC(x)_{OCT\_EXP}$. The motivation driving the definition of this term is analogous to that described for $CLOC(x)_{H\_dival}$ in section III.E.1. Specifically, we assume that there is overestimation of the nondynamical electron correlation for environments with overall higher electron density from neighboring orbitals. These systems are exemplified by the systems ClF_3 and PCl_5 where the central atom has a valence shell expansion beyond the usual octet. This term is also applied to transition states exemplified by the S_N2 reaction $F^- + CH_3Cl \rightarrow Cl^- + CH_3F$. Here, the carbon of the transition state also experiences an increase in electron density that also leads to an overestimation of nondynamical electron correlation. The overbinding of hypervalent structures is manifested in both atomization energies of hypervalent species and in transition states with hypervalent character, as is shown in detail in previous works such as refs 5a and 5e.

This LOC, $LOC(x)_{OCT\_EXP}$, is defined as

$$LOC(x)_{OCT\_EXP} = N_{OCT\_EXP} \times C_{OCT\_EXP} \tag{53}$$

where N_{OCT_EXP}, analogously to N_{H_dival} in eq 49, is

$$N_{OCT\_EXP} = \sum_i \eta_i \tag{54}$$

Here, i is an index that runs over all non-hydrogen atoms, and η_i is given by

$$\eta_i = \begin{cases} 0, & \text{if } n \leq 8 - g; \\ 1, & \text{if } n \geq (8 - g) + 1 \end{cases} \tag{55}$$

In this formula, n is the number of bonds around an atom center i, g is the element's group number on the periodic table, and "bonds" are defined as for eq 50. This equation ensures that atoms bonded to a number of elements that violate their octets are assigned $N_{OCT\_EXP} = 1$, whereas the opposite is true for atoms with a number of bonds that are within their octet.

For example, consider how chlorine is treated in HCl vs ClF_3. Chlorine's group number in the periodic table, g, is 7. In HCl, the number of bonds, n, to chlorine is one, and from the equation above, we have $\eta = 0$. Therefore, $N_{OCT\_EXP} = LOC(x)_{OCT\_EXP} = 0$; i.e., $LOC(x)_{OCT\_EXP}$ is not assigned for HCl. In ClF_3, however, we have $n = 3$, and from the equation above, $\eta = 1$. Therefore, $N_{OCT\_EXP} = LOC(x)_{OCT\_EXP} = 1$; i.e., $LOC(x)_{OCT\_EXP}$ is assigned for the chlorine of ClF_3.

We use these equilibrium values of N_{OCT_EXP} given by the equations above to determine $CLOC_{OCT\_EXP}$. Specifically, $CLOC_{OCT\_EXP}$ takes on the same general form as LOC_{OCT_EXP} in eq 53, except that ζ, which is continuous, is now used in place of N_{OCT_EXP}, which is discrete.

$$CLOC(x)_{OCT\_EXP} = \zeta \times C_{OCT\_EXP} \tag{56}$$

where

$$\zeta = \sum_i [f(N^{eq}_{OCT\_EXP}, N^{ts}_{OCT\_EXP})]_i \tag{57}$$

and i is an index that runs over all non-hydrogen atoms, i.e., those atoms which are eligible to receive $CLOC(x)_{OCT\_EXP}$. We define $f(N^{eq}_{OCT\_EXP}, N^{ts}_{OCT\_EXP})$ as a function of the equilibrium and transition state N_{OCT_EXP} values, $N^{eq}_{OCT\_EXP}$ and $N^{ts}_{OCT\_EXP}$, respectively. In this manner, the interpolated value of $CLOC(x)_{OCT\_EXP}$ for any intermediate structure x is a function of the stationary states' LOCs, $LOC(eq)_{OCT\_EXP}$ and $LOC(ts)_{OCT\_EXP}$. The equilibrium structure is taken as the reactant for a reactant-side arbitrary structure, or product for a product-side arbitrary structure.

$$f(N^{eq}_{OCT\_EXP}, N^{ts}_{OCT\_EXP}) = \begin{cases} 0, & \text{for } N^{eq}_{OCT\_EXP} = N^{ts}_{OCT\_EXP} = 0; \\ 1, & \text{for } N^{eq}_{OCT\_EXP} = N^{ts}_{OCT\_EXP} = 1; \\ f(\delta_x)_{hyb}, & \text{for } N^{eq}_{OCT\_EXP} = 0 \text{ and } N^{ts}_{OCT\_EXP} = 1; \\ 1 - f(\delta_x)_{hyb}, & \text{for } N^{eq}_{OCT\_EXP} = 1 \text{ and } N^{ts}_{OCT\_EXP} = 0 \end{cases} \tag{58}$$

where $f(\delta_x)_{hyb}$ is as defined in eq 30. As discussed earlier, $f(\delta_x)_{hyb}$ takes into account the multiple-bond-length dependency of hybridization. Because $CLOC(x)_{OCT\_EXP}$ also depends upon multiple bond lengths, $f(\delta_x)_{hyb}$ is employed here.

Inspection of this equation shows that where both the equilibrium structure (reactant for a reactant-side interpolation, or product for a product-side interpolation) and transition state do not receive an OCT_EXP correction, $N^{eq}_{OCT\_EXP} = N^{ts}_{OCT\_EXP} = 0$, the interpolated structure, $f(N^{eq}_{OCT\_EXP},$

$N^{ts}_{OCT\_EXP}) = 0$, does not either. This would apply to the reaction $CH_3\cdot + CH_2CH_2 \rightarrow CH_3CH_2CH_2\cdot$ depicted in Scheme 1, for example. The same holds for the opposite case. Namely, where both the equilibrium structure and transition state structure do receive $LOC(x)_{OCT\_EXP}$, so does the interpolated structure. This would apply to the reaction $SO_4^{2-} + H_3O^+ \rightarrow HSO_4^- + H_2O$, for example, where the sulfur atom merits $LOC(x)_{OCT\_EXP}$ throughout the reaction. For a reaction where the equilibrium structures do not merit $LOC(x)_{OCT\_EXP}$, yet the transition state does, the amount of $CLOC(x)_{OCT\_EXP}$ the interpolated structure receives is proportional to $f(\delta_x)_{hyb}$. Therefore, the amount of $CLOC(x)_{OCT\_EXP}$ increases smoothly toward the transition-state value as the structure itself becomes more transition-state-like, as quantified by $f(\delta_x)_{hyb}$, defined by eq 30 above. This applies to the reaction $FCH_3 + Cl^- \rightarrow F^- + CH_3Cl$, for example, where neither equilibrium structure (reactant or product) merits $LOC(x)_{OCT\_EXP}$, yet the transition state does. Lastly, where the equilibrium structure does merit $LOC(x)_{OCT\_EXP}$ but the transition state does not, the amount of $CLOC(x)_{OCT\_EXP}$ decreases smoothly toward the transition-state value, again as a function of $f(\delta_x)_{hyb}$.

None of the reactions in Scheme 2 merit the $CLOC(x)_{OCT\_EXP}$ correction, and hence this specific term has not yet been tested. However, we anticipate that this correction will work well judging from the behavior of all of the other similar terms.

III.F. CLOCs for Environment, $CLOC(x)_{environ}$. We have also argued that the presence of neighboring bonds connected to a particular base bond contributes to systematic error in the quantification of nondynamical correlation of that base bond,[5] and we introduce the $LOC(x)_{environ}$ term, $LOC(x)_{ESBC}$, to capture these effects. This term arises from the fact that an electron in the base bond can make an excursion to a neighboring bond, increasing its nondynamical correlation energy, particularly if it is a long bond, vs a single bond to hydrogen, for example.

We stated in section III.A that it is necessary that $CLOC(x)$ agrees with $LOC(x)$ where x is a stationary point (reactant, product, or transition state). To meet this requirement, we found it necessary to slightly modify the way the previously defined $LOC(x)_{ESBC}$ parameter was extended to transition states in the latest LOC publication.[5e] Let us begin with a detailed explanation of how $LOC(x)_{ESBC}$ was assigned to transition states in the previous work[5e] to understand why a modification was necessary.

While the definition of $LOC(x)_{ESBC}$ for reactants and products is straightforward, formulating an implementation for transition states is less obvious. For example, consider the Diels−Alder reaction between butadiene and ethene (reaction a in Scheme 2). The reactant and product are readily assigned $N_{ESBC} = 0$ and 8, respectively. Yet, the $LOC(x)_{ESBC}$ assignment for the transition state is not immediately obvious.

Because $LOC(x)_{ESBC}$ is applied only for neighboring single bonds, the bond order s_i for each bond is transformed into a value to describe its percent single bond character, $f(s_i)$, on a scale from 0−1; 1 being a pure single bond and 0 being no bond or a pure double bond.

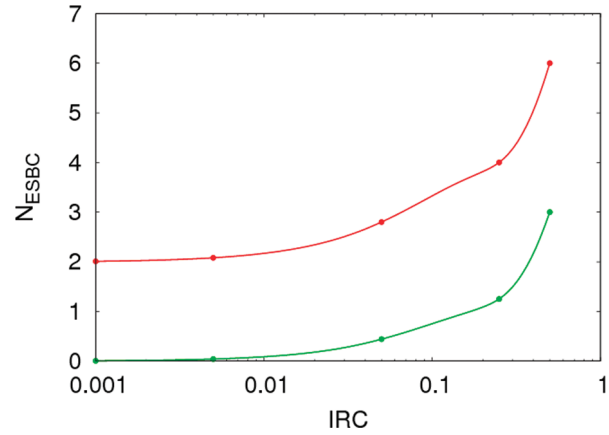


Figure 5. N_{ESBC} vs intrinsic reaction coordinate (IRC) for reactant to transition state of reaction a in Scheme 2. N_{ESBC} as defined in eqs 59−61 is shown in red, while N_{ESBC} as defined by eqs 59 and 62−64 is shown in green.

$$f(s_i) = e^{-\gamma(s_i-1)^2} \tag{59}$$

Here, γ here is chosen to be 3 such that $f(s_i) \approx 0.5$ for $s_i = 0.5$.

The N_{ESBC} for any bond i, $(N_{ESBC})_i$, is then given by the sum of $f(s_j)$ over all neighboring bonds j.

$$(N_{ESBC})_i = \sum_j f(s_j) \tag{60}$$

and the total N_{ESBC} for the system is then the sum over all $(N_{ESBC})_i$ for each bond i.

$$N_{ESBC} = \sum_i (N_{ESBC})_i \tag{61}$$

Using the above formulas, we find that the reactant and transition state of reaction a in Scheme 2 have $N_{ESBC} = 0$ and 6, respectively. In Figure 5, we show the results of using the formulations given in eqs 59−61 to interpolate points intermediate between the reactant and transition state for this reaction in red. Importantly, we see that $N_{ESBC} \rightarrow 2$ as the reaction coordinate $\rightarrow 0$. Recall that for the reactant, $N_{ESBC} = 0$, and therefore, $CLOC(x)$ does not agree with $LOC(x)$ as we approach the reactant. Yet, we stated in section III.A that $CLOC(x)$ must agree with $LOC(x)$ where x is a stationary point.

To force agreement between $CLOC(x)$ and $LOC(x)$ at the reactant and product, we have changed the formulation of $LOC(x)_{ESBC}$ for transition states from that described by eqs 59−61. In the previous publication,[5e] it was only the neighboring bonds and their bond orders, s_j, which determined N_{ESBC} for the bond under consideration, but not the bond order of that bond itself, s_i (see eq 60). In the latest implementation, both the bond order, s_i, of the bond under consideration and that of its neighboring bond, s_j, are considered when determining N_{ESBC}. Specifically, both $f(s_i)$ and $f(s_j)$, defined in eq 59, are multiplied to give a number that reflects cumulative percent single bond character for the pair, π_{ij}

$$\pi_{ij} = f(s_i) \times f(s_j) \tag{62}$$

In this manner, two neighboring single bonds receive the maximum value ($\pi_{ij} = 1$), while a neighboring single bond and "half bond" will receive a lesser value ($\pi_{ij} = 1/2$), and two neighboring "half" bonds will receive a lesser value still ($\pi_{ij} = 1/4$), for example. This reflects the fact that the more single-bond-like the neighboring bonds are, the more excursions are possible from a base bond into these neighboring bonds, and hence the more correction is necessary to account for these excursions.

For each bond i, the sum of all π_{ij} values is taken across all neighboring bonds j, which produces the total N_{ESBC} for that bond, $N_{\mathrm{ESBC}}(\pi_{ij})$.

$$N_{\mathrm{ESBC}}(\pi_{Ij}) = \sum_{i<j} (\pi_{ij}) \tag{63}$$

The increase in π_{ij} with increasing single-bond-character is thus utilized here to also assign larger $N_{\mathrm{ESBC}}(\pi_{ij})$ for systems with more neighboring bonds with high single-bond character.

All $N_{\mathrm{ESBC}}(\pi_{ij})$'s for each unique ij pair are summed to give the final N_{ESBC} for that system.

$$N_{\mathrm{ESBC}} = \sum_{i<j} N_{\mathrm{ESBC}}(\pi_{ij}) \tag{64}$$

The results of the new definition of N_{ESBC}, as given in eqs 59 and 62−64, are shown in Figure 5. Notice that while the previous definition of N_{ESBC} (shown in red in Figure 5) did not have the proper behavior, i.e., N_{ESBC} did not approach zero as the reaction coordinate → 0, this definition of N_{ESBC} (shown in green in Figure 5) does indeed have the proper behavior, i.e., $N_{\mathrm{ESBC}} \rightarrow 0$ as the reaction coordinate → 0. Therefore, we have satisfied the requirement that CLOC(x) agree with LOC(x) where x is a stationary point, at least for x being the reactant or product.

Inspection of Figure 5 also shows that N_{ESBC} for the transition state of this reaction has changed, from $N_{\mathrm{ESBC}} = 6$ in the old definition to $N_{\mathrm{ESBC}} = 3$ in the new definition. As with any parametrization, we are free to change the definition of how parameters are applied (N_k in eq 1) so long as we reoptimize the values of the parameters (C_k in eq 1) in accordance with their new definitions. Therefore, the newly defined application of N_{ESBC} to transition states necessitated reoptimizing the values of the transition-state specific parameters, C_k, to optimally reduce the B3LYP error in barrier heights. The updated values, which are only slightly different than those previously published, and all LOC parameter values and definitions can be found in the Supporting Information. Note that while the individual B3LYP-LOC barrier heights have changed slightly, the overall performance of B3LYP-LOC remains unchanged. That is, the LOCs still produce a dramatic reduction in the B3LYP barrier height errors and predict barrier heights within or near chemical accuracy (traditionally taken as ≤1 kcal/mol) across a broad spectrum of reactions.

III.G. CLOC for Charge Transfer, CLOC(x)$_{\mathrm{CT}}$. In section III.B, we argue that as the bond length increases, nondynamical correlation becomes more negative (as the electrons have more room to avoid each other), and DFT systematically underestimates this effect with increasing severity. An extreme example of this is manifest in systems such as carbon monoxide, $^{-}C{\equiv}O^{+}$, or sodium chloride, $Na^{+}Cl^{-}$, which have zero overall formal charge but nonzero formal charge on individual atoms. In these systems, the localized orbitals are highly ionic in character and hence compactly organized around one of the two atoms, and the bonds are also relatively long in comparison to the size of the orbitals in which the electron pairs are localized. Because this is a severe example of underbinding, this situation when it arises receives its own special parameter, LOC(x)$_{\mathrm{CT}}$, according to

$$\mathrm{LOC}(x)_{\mathrm{CT}} = N_{\mathrm{CT}} \times C_{\mathrm{CT}} \tag{65}$$

where C_{CT} is the optimized value of the LOC(x)$_{\mathrm{CT}}$ parameter (see the Supporting Information) and N_{CT} is given by

$$N_{\mathrm{CT}} = \sum_{i<j} f(q_i) \times f(q_j) \tag{66}$$

Here, i and j are indices that run over all neighboring atom pairs and $f(q)$ is a function of the formal charge q on an atom (as defined in eq 31) given by

$$f(q) = \begin{cases} 0, & \text{for } q = 0; \\ 1, & \text{for } |q| \geq 1 \end{cases} \tag{67}$$

An inspection of this equation shows that N_{CT} and LOC(x)$_{\mathrm{CT}}$, by extension, are nonzero only where two neighboring atoms both have nonzero charge.

The continuous version of LOC(x)$_{\mathrm{CT}}$, CLOC(x)$_{\mathrm{CT}}$, may be written analogously as

$$\mathrm{CLOC}(x)_{\mathrm{CT}} = N_{\mathrm{CT}} \times C_{\mathrm{CT}} \tag{68}$$

where N_{CT} is given still by eq 66 and only the definition of $f(q)$ is modified to allow for continuous representation of partial formal charges.

$$f(q) = \begin{cases} 0, & \text{for } q = 0; \\ e^{-\gamma(|q| - 1)^2}, & \text{for } 0 < |q| < 1; \\ 1, & \text{for } |q| \geq 1 \end{cases} \tag{69}$$

where γ, as before, is chosen to be 3.0 such that $f(|q|) \approx 0.5$ for $|q| = 0.5$. This continuous definition of $f(q)$ is identical to the former discrete version with the exception that it allows for noninteger charges on atoms. Inspection of this equation further shows that as the absolute values of charges on any two neighboring atoms approach 1, $|q| \rightarrow 1$, then $f(q) \rightarrow 1$, and thus $N_{\mathrm{CT}} \rightarrow 1$, its maximal value. Therefore, we are equipped to treat noninteger partial charges in a smooth and continuous fashion and give only the maximal value of N_{CT} to systems with integer values of formal charge on neighboring atoms.

Because none of the reactions in Scheme 2 merit the CLOC(x)$_{\mathrm{CT}}$ correction, this specific term has not yet been tested. However, we anticipate its correct behavior on account of the behavior of all of the other similar terms.

III.H. Total CLOC(x). As stated in section III.A, the total CLOC(x) is given by the sum of its constituents:

$$\mathrm{CLOC}(x) = \mathrm{CLOC}(x)_{\mathrm{bond}} + \mathrm{CLOC}(x)_{\mathrm{hyb}} + \mathrm{CLOC}(x)_{\mathrm{radical}} + \mathrm{CLOC}(x)_{\mathrm{hyperval}} + \mathrm{CLOC}(x)_{\mathrm{environ}} + \mathrm{CLOC}(x)_{\mathrm{CT}} \quad (70)$$

Therefore, to arrive at the total CLOC(x) for any arbitrary x, the individual components of this expression are calculated according to the prescriptions given in sections III.B−III.G, and summed over. This CLOC(x) may then be used directly to obtain more accurate relative energies, as described in section III.A.

Once CLOC(x) is known for any arbitrary x, we may define gradients of the B3LYP-CLOC functional. They are given according to the formula

$$\nabla E_{\mathrm{B3LYP-CLOC}}(x) = \nabla[E_{\mathrm{B3LYP}}(x) + \mathrm{CLOC}(x)] = \nabla E_{\mathrm{B3LYP}}(x) + \nabla\mathrm{CLOC}(x) \quad (71)$$

III.I. Computational Methods. All intrinsic reaction coordinate (IRC) scans were performed at the B3LYP/6-31+G** level using the computational package Jaguar 7.6.[9] (In previous LOC publications,[5] all geometry optimizations and transition state searches were performed at the B3LYP/6-31G* level; however, we found that using this slightly larger basis greatly improved the ease with which stationary points could be located, without substantially increasing computational cost.) These geometries were then used to perform single-point energy calculations at the B3LYP/6-311++G(3df,3pd) and M06-2X/6-311++G(3df,3pd) levels, also within Jaguar; at the RCCSD(T)/cc-pVTZ level using MolPro 2006.1;[8] and at the BW2 post-HF level using the code provided by Hans Joachim-Werner for reaction g.[14] Following Joachim-Werner's precedent for this reaction, a mixed basis was used in which chlorine was treated with the aug-cc-pV5Z[8s7p5d4f3g] basis, and hydrogens were treated with the aug-cc-pVQZ[5s4p3d2f] basis.

The potential energy curves obtained with the post-HF, B3LYP, M06-2X, and B3LYP-CLOC methods were aligned by relative energy. Specifically, the energy of the product(s) was subtracted from the energy at every point along the curve such that the energy of the product(s) for all four curves was strictly zero, and all other energies were given with respect to the product(s) energy.

The B3LYP-CLOC curves were generated directly from the B3LYP curves with the simple addition of the numerical CLOC. For example, to compute the B3LYP-CLOC energy for an arbitrary point x on the reaction profile, $E_{\mathrm{B3LYP\text{-}CLOC}}(x)$, the CLOC at point x, CLOC(x), had to be initially obtained. The B3LYP-CLOC energy at point x is then the sum of the B3LYP energy and CLOC.

$$E_{\mathrm{B3LYP}}(x) + \mathrm{CLOC}(x) = E_{\mathrm{B3LYP-CLOC}}(x) \quad (72)$$

Since all reaction profile curves have to be scaled by the subtraction of the product(s) energy, we must know the B3LYP-CLOC energy of the product(s). This is computed similarly to the description given above.

$$E_{\mathrm{B3LYP}}(\mathrm{product}) + \mathrm{CLOC}(\mathrm{product}) = E_{\mathrm{B3LYP-CLOC}}(\mathrm{product}) \quad (73)$$

Finally, the energies along the B3LYP-CLOC curves are given by the difference in $E_{\mathrm{B3LYP\text{-}CLOC}}(x)$ and $E_{\mathrm{B3LYP\text{-}CLOC}}$(product).

$$E^{\mathrm{relative}}_{\mathrm{B3LYP-CLOC}}(x) = E_{\mathrm{B3LYP-CLOC}}(x) - E_{\mathrm{B3LYP-CLOC}}(\mathrm{product}) \quad (74)$$

These are the final points given in all of the graphs and tables in this work.

All numerical CLOCs can be computed with a simple script.[15] As input, a reaction coordinate (defined by reactant, transition state, and product structures) and an arbitrary structure along that coordinate are required, and CLOCs for all four structures are produced in the output.

IV. Results and Discussion

To assess the effectiveness of the CLOC approach, we surveyed reaction profiles for the seven reactions shown in Scheme 2. Three functionals—B3LYP, M06-2X, and B3LYP-CLOC developed herein—were tested against post-HF level calculations, with the resulting plots shown in Figure 6. Table 6 and Figure 7 show the mean unsigned errors (MUEs) along each reaction profile for all reactions and functionals studied. These numbers reflect the disagreement between post-HF and DFT at every point along the reaction profile. The mean value of all of these MUEs, MMUE(overall), is also given.

While MMUE(overall) reflects the performance across the entire reaction coordinate, the performance at the stationary points for most practical applications is more critical than the performance at intermediate points. Therefore, the differences in relative SCF energies at the stationary points, ΔE_{scf}, are compared to the values at the post-HF level in Table 7. Specifically, the differences in transition state and equilibrium SCF energies, ΔE_{scf}(eq→ts), where the equilibrium structure may be either reactant or product, were tabulated in addition to the differences in reactant and product SCF energies, ΔE_{scf}(r→p). The mean unsigned errors in these two values, ΔE_{scf}(eq→ts) and ΔE_{scf}(r→p), across all seven reactions were then tabulated for each DFT method to give the final values shown in Table 7, MUE[ΔE_{scf}(eq→ts)] and MUE[ΔE_{scf}(r→p)].

An inspection of the plots in Figure 6 not surprisingly shows that while the B3LYP curves have the qualitatively correct shape compared to the post-HF standard in many cases, there still remain large quantitative errors at many points along the reaction coordinate. This is also reflected by the relatively large values of MUE[ΔE_{scf}(eq→ts)] and MUE[ΔE_{scf}(r→p)] given in Table 7. Perhaps fortuitously, there are some reactions for which the B3LYP and post-HF curves are nearly convergent, at least for part of the reaction profile, i.e., reaction d, CH_3· + CH_2CH_2. However, there are other curves where serious quantitative disagreement between B3LYP and the post-HF method is observed, i.e., reaction g, H_2+Cl.

Further inspection of Figures 6 and 7 and Tables 6 and 7 shows that, on average, M06-2X outperforms B3LYP across all stationary and intermediate structures. This is reflected by lower values of MUE[ΔE_{scf}(eq→ts)] and MUE[ΔE_{scf}-(r→p)] in Table 7 and MMUE(overall) in Table 6 for M06-2X vs B3LYP. Yet, we see that B3LYP performs as well

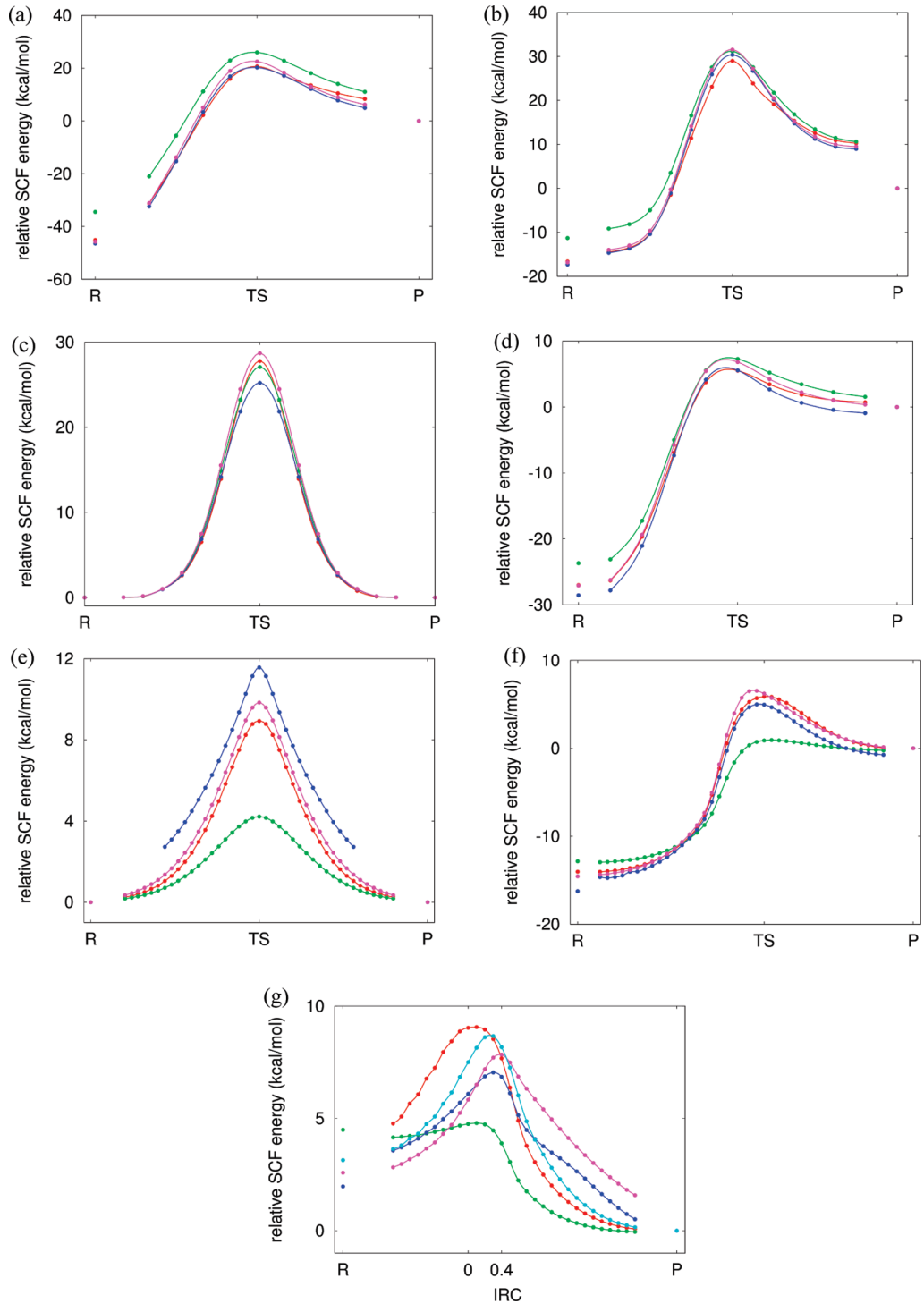


Figure 6. Reaction profile plots for the reactions of Scheme 2. All plots show relative SCF energies with respect to intrinsic reaction coordinate (IRC), where the points on the coordinate are labeled R, TS, and P for reactant, transition state, and product, respectively. The plots are constructed for the following reactions: (a) Diels−Alder, (b) electrocyclic, (c) sigmatropic shift, (d) carbon radical reaction, and hydrogen radical reactions (e) $H_2 + H$, (f) $H_2 + OH$, and (g) $H_2 + Cl$. The curves plotted are B3LYP (green), B3LYP-CLOC (red), post-HF (magenta), and M06-2X (blue). For reaction g, B3LYP-CLOC with a shifted IRC for the transition state is also plotted in teal. Note that in this plot, the B3LYP-optimized transition state occurs at IRC = 0.0, while the post-HF transition state occurs at IRC = 0.4. Post-HF plots for reactions a−f are at the RCCSD(T) level, while the post-HF plot for reaction g is at the BW2 level.

as, or better than, M06-2X for both the sigmatropic shift (reaction c) and carbon radical reaction (reaction d) in Table 6 when we consider the MUE along the entire reaction coordinate.

Similarly to M06-2X, B3LYP-CLOC performs better than B3LYP along the whole reaction coordinate for many different reaction types. The same trend holds for both ΔE_{scf}(eq→ts) and ΔE_{scf}(r→p). [Notably, however, all func-

Table 6. Mean Unsigned Errors (MUEs)[a] in kcal/mol along Entire Reaction Profile

reaction	reaction type	B3LYP	M06-2X	B3LYP-CLOC
a	Diels−Alder	5.69	1.27	1.39
b	electrocyclic	2.30	0.62	1.31
c	sigmatropic shift	1.33	1.35	0.57
d	carbon radical	0.36	0.78	0.56
e	$H_2 + H$	2.66[b]	1.43[b]	0.55[b]
f	$H_2 + OH$	2.12	0.89	0.34
g	$H_2 + Cl$	2.34	0.93	2.28 (1.41[c])
MMUE(overall)[d]		2.40	1.04	1.00 (0.88[c])

[a] The deviation between post-HF and DFT energies at every point along the reaction coordinate was computed. The absolute values of these deviations were then averaged to give the mean unsigned error, MUE, along the entire curve. [b] All data computed using only data along the range $-1.4 \leq$ IRC ≤ 1.4 because M06-2X data points outside this range could not be obtained due to convergence difficulties. [c] Computed using the shifted B3LYP-CLOC data for the reaction $H_2 + Cl$. [d] Mean of MUEs for reactions *a*−*g*, i.e., the values given in the rows directly above.

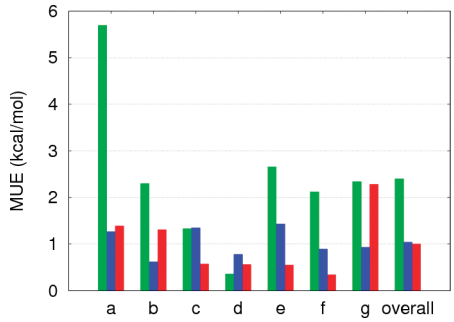


Figure 7. Mean unsigned error (MUE) vs reaction type for all reactions and functionals tested in this study as shown in Table 6. Functionals include B3LYP (green), M06-2X (blue), and B3LYP-CLOC (red). Reactions include (a) Diels−Alder, (b) electrocyclic, (c) sigmatropic shift, (d) carbon radical, (e) $H_2 + H$, (f) $H_2 + OH$, and (g) $H_2 + Cl$.

Table 7. Mean Unsigned Errors (MUEs) in kcal/mol for ΔE_{scf}(eq→ts) and ΔE_{scf}(r→p)

	B3LYP	M06-2X	B3LYP-CLOC
MUE[ΔE_{scf}(eq→ts)][a]	4.1	1.3	1.5
MUE[ΔE_{scf}(r→p)][b]	3.4	0.7	0.3

[a] Mean unsigned error of ΔE_{scf}(eq→ts), where eq may be either equilibrium structure, reactant(s), or product(s), for reactions a−g. Individual errors for ΔE_{scf}(eq→ts) given in Supporting Information. All ΔE_{scf}(eq→ts) were computed using B3LYP geometries. [b] Mean unsigned error of ΔE_{scf}(r→p) for reactions a−g. Individual errors for ΔE_{scf}(r→p) given in Supporting Information. All ΔE_{scf}(r→p) were computed using B3LYP geometries.

tionals tested have difficulty with reaction g, $H_2 + Cl$. This will be discussed further below.] While M06-2X is outperformed by B3LYP for both reactions c and d, B3LYP-CLOC is only outperformed by B3LYP in the case of reaction d. Arguably, this is a case in which B3LYP performs anomalously well. An inspection of Table 6 shows that B3LYP's performance for reaction d is far better than for all the other reactions studied. Additionally, examination of Tables 6 and 7 and Figure 7 reveals that the performance of B3LYP-CLOC rivals that of M06-2X with respect to MUE for the individual cases as well as all cases combined. This is also true for the MUE[ΔE_{scf}(eq→ts)] and MUE[ΔE_{scf}(r→p)], as shown in Table 7.

Table 8. Geometry of $H_2 + Cl$ Transition State

	bond length (Å)	
method	H−H	H−Cl
BW2[a]	1.00	1.46
B3LYP	1.29	1.34
B3LYP-CLOC	1.29	1.34

[a] Taken as the geometry at IRC = 0.4 in Figure 6g.

Overall, the performance of B3LYP-CLOC rivals that of M06-2X. Both methods exhibit similar accuracy over the entire test set, but M06-2X performs appreciably better than B3LYP-CLOC for reactions b and g, whereas B3LYP-CLOC performs appreciably better than M06-2X for reactions c, e, and f. More reactions in addition to those examined in this work will have to be studied before a broad conclusion about the relative performance of these two functionals may be drawn.

Reaction g, as stated previously, is particularly problematic. This exception may be explained at least in part by the fact that the B3LYP and post-HF transition states' geometries differ markedly from one another, as shown in Table 8. In fact, an inspection of the reaction profile in Figure 6g shows that the post-HF transition state actually occurs at IRC = 0.4, where the geometry more closely matches that found by the post-HF transition state search. Interestingly, the same shift of transition state along the reaction coordinate is observed for M06-2X, where the transition state occurs at IRC = 0.3. We observe that all three tested DFT methods—B3LYP, B3LYP-CLOC, and M06-2X—have difficulty in accurately predicting the transition state geometry for this complicated case, where notably M06-2X performs better than B3LYP and B3LYP-CLOC.

Since our algorithm requires reactant, product, and transition state structures as input to handle all intermediate points along the reaction coordinate, and we know that the "true" transition state geometry resembles that at IRC = 0.4, we can instead provide this alternate geometry as the input transition state geometry to our algorithm and therefore shift the location of the transition state on the reaction coordinate accordingly. This is indeed what we have done to produce the teal curve in Figure 6g. Note that all points along the reaction profile now better approximate the post-HF curve. A complete solution to this problem would not require prior knowledge of the post-HF geometry. Specifically, we ultimately seek a method whereby we can produce energy curves as accurate as post-HF methods without prior knowledge of the energy curves or geometries produced by these methods whatsoever.

Disagreement between B3LYP and "true" transition-state geometries is not without precedent. This same behavior is observed for the highly analogous reaction $H_2 + F$, where transition states predicted by B3LYP and post-HF methods differ substantially in bond lengths and angles.[15]

The above analysis suggests that the inability of DFT to treat both these problematic cases stems partly from defective reproduction of geometries predicted with high-level post-HF methods. One possible solution is the alteration of DFT functionals in such a way that they produce geometries more closely matching the post-HF ones, presumably also leading to more accurate energies. This can be achieved via alteration of

the DFT gradients, ∇E_{DFT}. As stated in section III.H, gradients for the B3LYP-CLOC functional are given according to

$$\nabla E_{B3LYP-CLOC}(x) = \nabla[E_{B3LYP}(x) + CLOC(x)] = \nabla E_{B3LYP}(x) + \nabla CLOC(x) \quad (75)$$

In this regard, it is important to consider the relative magnitudes of the two terms on the right-hand side of eq 75. Specifically, $\nabla E_{B3LYP}(x)$ is often much larger than $\nabla CLOC(x)$, which is only ever a few killocalories per mole, such that $\nabla E_{B3LYP\text{-}CLOC}(x) \approx \nabla E_{B3LYP}(x)$. In fact, we see no change whatsoever in the transition state geometries, as $\nabla E_{B3LYP}(x) = \nabla E_{B3LYP\text{-}CLOC}(x)$. This is clear upon inspection of Table 8, where we see that the geometry of the H_2 + Cl transition state is identical for both B3LYP and B3LYP-CLOC. For non-transition-state structures, we expect only very small changes in the B3LYP-CLOC geometries vs B3LYP.

Therefore, the power of the B3LYP-CLOC method does not lie in its ability to produce more accurate geometries, but rather in its ability to produce more accurate energies. This is most useful for reactions where B3LYP already produces reasonably accurate geometries. Fortunately, most reactions of practical interest fall into this category. For example, note that the larger systems employed in this study, such as reactions a−d in Scheme 2, do not suffer from the difficulties encountered with H_2 + Cl and H_2 + F. A particularly attractive characteristic of our method is therefore its ability to deliver highly accurate, yet computationally inexpensive energies for larger systems.

V. Conclusions

In this work, we have shown how simple empirical localized orbital corrections (LOCs) can be generalized to formulate a continuous implementation (CLOC) that is defined throughout a reaction profile. These corrections were applied specifically to the B3LYP functional, as this functional has shown itself most amenable to this correction scheme. The resultant method, B3LYP-CLOC, gives more accurate energetics in comparison to B3LYP, and its accuracy rivals that of M06-2X for the test cases examined. Furthermore, negligible additional computational cost is required over standard B3LYP calculations, and convergence of geometry optimizations is facile. The accuracy is best where B3LYP already produces reasonable geometries and assignment of Lewis structures is straightforward.

Future work will focus on extending this implementation to the treatment of ionic reactions in addition to the neutral reactions studied herein. More reaction profiles should be studied to test the robustness of this method.

Acknowledgment. This work was supported in part by a grant from the National Institute of Health (NIH) training program in Molecular Biophysics (M. L. Hall, T32GM008281) and by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy through Grant DE-FGO2-903R14162.

Supporting Information Available: Cartesian coordinates for all structures along with their relative energies at the post-HF, B3LYP, M06-2X, and B3LYP-CLOC levels are provided. The complete suite of all LOCs in addition to detailed examples of parameter assignments and their application to a database of over 105 barrier heights is also available. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kohn, W.; Becke, A. D.; Parr, P. G. *J. Phys. Chem.* **1996**, *100*, 12974–12980.

(2) For a review of DFT, the history of functional development, and an assessment of the performances of various popular functionals, see: Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A.* **2007**, *111*, 10439–10452.

(3) (a) Oliveira, G.; Martin, J. M. L.; Proft, F.; Geerlings, P. *Phys. Rev. A* **1999**, *60*, 1034–1045. (b) Izgorodina, E. I.; Brittain, D. R. B.; Hodgson, J. L.; Krenske, E. H.; Lin, C. Y.; Namazian, M.; Coote, M. L. *J. Phys. Chem. A* **2007**, *111*, 10754–10768. (c) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2001**, *105*, 2936–2941.

(4) (a) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103. (b) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A.* **2000**, *104*, 4811–4815. (c) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A.* **2004**, *108*, 6908–6918.

(5) (a) Friesner, R. A.; Knoll, E. H.; Cao, Y. J. *J. Chem. Phys.* **2006**, *125*, 124107. (b) Knoll, E. H.; Friesner, R. A. *J. Phys. Chem. B* **2006**, *110*, 18787–18802. (c) Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Phys.* **2008**, *129*, 214105. (d) Rinaldo, D.; Tian, L.; Harvey, N. J.; Friesner, R. A. *J. Chem. Phys.* **2008**, *129*, 164108. (e) Hall, M. L.; Goldfeld, D. A.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Theory Comput.* **2009**, *5*, 2996–3009.

(6) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(7) (a) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2004**, *121*, 8187. (b) Polo, V.; Kraka, E.; Cremer, D. *Mol. Phys.* **2002**, *100*, 1771–1790. (c) Grafenstein, J.; Kraka, E.; Cremer, D. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1096–1113. (d) Grafenstein, J.; Kraka, E.; Cremer, D. *J. Chem. Phys.* **2004**, *120*, 524. (e) Cremer, D. *Mol. Phys.* **2001**, *99*, 1899–1940. (f) Polo, V.; Kraka, E.; Cremer, D. *Mol. Phys.* **2002**, *100*, 1771–1790.

(8) *MOLPRO*, version 2006.1; MOLPRO: Cardiff, U.K., 2006.

(9) *Jaguar*, version 7.6; Schrödinger, LLC: New York, 2009.

(10) Bian, W. S.; Werner, H. J. *J. Chem. Phys.* **2000**, *112*, 220.

(11) For reviews of the application of DFT to the study of organic chemistry and biochemistry, see: (a) Roos, G.; Geerlings, P.; Messens, J. *J. Phys. Chem. B.* **2009**, *113*, 13465–13475. (b) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. *J. Chem. Theory Comput.* **2007**, *3*, 407–433.

(12) Zhang, Y. K.; Yang, W. T. *J. Chem. Phys.* **1998**, *109*, 2604.

(13) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(14) Joachim-Werner, H. The $Cl + H_2 \rightarrow HCl + H$ Reaction. http://www.theochem.uni-stuttgart.de/~werner/h2cl/h2cl.html (accessed Jun 1, 2010).

(15) Werner, H.-J.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2008**, *128*, 034035.

CT100418N