

ARTICLE

Efficient and Flexible Approach for Local Distortion: Distortion Distribution Analysis enabled by Fragmentation

Received 00th January 20xx,
Accepted 00th January 20xx

Zeyin YAN,^{†a} Yunteng Sam Liao,^{†a} Xin Li^a and Lung Wa Chung^{*a}

DOI: 10.1039/x0xx00000x

Distortion can play crucial roles in influencing structures and properties, as well as enhancing reactivity or selectivity in many chemical and biological systems. The distortion/interaction model is a popular and powerful method for deciphering the origins of activation energies, in which distortion and interaction energies dictate an activation energy. However, decomposition of local distortion energy at the atomic scale remains less clear and straightforward. Knowing such information should deepen our understanding of reaction processes and improve reaction design. Herein, an efficient, general and flexible fragmentation-based approach was proposed to evaluate local distortion energies for various chemical and biological molecules. Moreover, our distortion analysis is readily applicable to multiple structures from molecular dynamics (or minimum energy path) as well as can be evaluated by different computational chemistry methods. Our systematic analysis shows that our approach not only visualizes (relative) distortion distributions within molecules (distortion map) and identifies the key distorted pieces, but also offers deeper understanding and insights into structures, reaction mechanisms and dynamics in various chemical and biological systems. Furthermore, our analysis offers indices of local distortion energy, which can potentially serve as a new descriptor in multi-linear regression (MLR) or machine learning (ML) modelling.

Introduction

Understanding chemical processes (such as reaction mechanism, kinetics, and selectivity) has been of paramount importance, helping chemists to rationally design catalysts and optimize reactions. Numerous theories and models have also been emerged or developed to provide valuable explanations and insights for understanding and improving reactions. For instance, pioneering theories, such as transition state theory (TST)¹ and Marcus theory,² have laid the vital groundwork for understanding chemical kinetics. Additionally, orbital-based theories, such as frontier molecular orbital (FMO),³ Woodward–Hoffmann rules,⁴ valence bond (VB),⁵ natural bond orbitals (NBO),⁶ intrinsic bond orbitals (IBO),⁷ and principal interacting orbitals (PIO) theories⁸ offer more intuitive insights and analysis of reactions from an orbital perspective.

Notably, deformation/interaction, distortion/interaction and activation-strain models pioneered by the Morokuma, Houk and Bickelhaupt groups,^{9–12} are popular and powerful methods for deciphering the origin of activation energies. This model decomposes an activation energy (or relative electronic energy along the reaction coordinate (ζ) or intrinsic reaction coordinate;¹³ ΔE^\ddagger or $\Delta E(\zeta)$, respectively) into distortion energy

(E_{dist}) of the reactant fragments (**A** and **B**) and their interaction energy (E_{int} , Scheme 1). This approach has been successfully and widely applied to various homogeneous and biological systems.^{14–15}

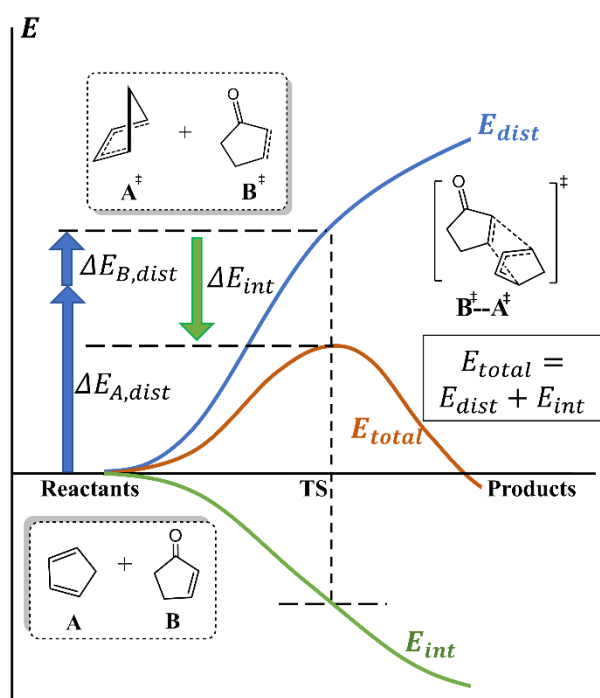
When the interaction energy contribution is crucial, various components of interactions can be further analyzed and quantified by energy decomposition analysis (EDA)¹⁶ and its variants.^{17–23} Additionally, to analyze, identify or correct (weak) noncovalent interactions (NCIs), diversified methods and tools (e.g., reduced density gradient,²⁴ dispersion interaction density/potential²⁵ and classical or nonlocal density-based dispersion²⁶) were also developed.^{27–28} Moreover, Fukui functions,²⁹ conceptual density functional theory,³⁰ quantum theory of atoms in molecules (QTAIM),³¹ electrostatic potential maps,³² steric maps³³ and Sterimol^{34–35} have been widely utilized to gain insights into electronic or steric effects on chemical systems.³⁶

Despite many method development and tools to analyze and decompose interaction energy, detailed analysis and tools for distortion energy (also known as deformation or strain energy³⁷) have received much less attention. It should be noted that distortion energy has played important roles in affecting some structures and properties as well as promoting some reactivity or selectivity.^{38–53} What's more, distortion energy is regarded as a global property of a molecule, where local atomic geometry should collectively determine the (global) distortion energy of the molecule. However, the distribution/decomposition of (local) distortion energy at the atomic scale is less clear or straightforward. Acquisition of such information should enhance our understanding on the reaction processes.

^aShenzhen Grubbs Institute, Department of Chemistry and Guangdong Provincial Key Laboratory of Catalysis, Southern University of Science and Technology, Shenzhen 518055, China. E-mail: oscarchung@sustech.edu.cn

[†]Electronic supplementary information (ESI) available.: Detailed method description; additional benchmarks; detailed results of systems in the main text and results of additional systems. See DOI: 10.1039/x0xx00000x

^{*}These authors contributed equally.



Scheme 1 Distortion/interaction analysis

To address this issue, the Dreuw group developed a seminal method (so-called JEDI) to evaluate and visualize distortion in mechanochemical systems, based on quantum-mechanics (QM)-computed Hessian matrix using redundant internal coordinates and harmonic approximation.⁵⁴ In addition, the Jasti group elegantly developed a strain-visualization method (StrainViz) for macrocycles specifically by combining isodesmic reaction, iterative fragmentation and integrating forces released into each internal coordinate during geometry optimization of each fragment.⁵⁵ These two methods have successfully been applied to a few specific systems and help illustrate the local distortion distribution.^{56–58} However, computational cost of hessian for medium and large systems is expensive, especially for some computational chemistry methods lacking analytic hessian.

To elucidate (relative) distortion distribution of various systems at different stages (e.g. stationary points, and/or any (non-stationary-point) structures from IRC or molecular dynamics (MD) simulations), an efficient, general and flexible fragmentation-based approach (Distortion Distribution Analysis enabled by Fragmentation, D2AF) was proposed to qualitatively estimate local distortion energies for various chemical and biological molecules. Our approach not only visualizes distortion energy distribution within molecules (i.e. distortion map) and identifies the key distorted part(s), but also provides values of local distortion energy, that can also be used as a new descriptor in statistical (e.g. multi-linear regression, MLR) or machine learning (ML) analysis/models.⁵⁹

Methodology

As outlined in Fig. 1, our fragmentation-based approach consists of three stages: fragmentation, calculations, and analysis. The

users first provide two structures: target molecule (**Tar**; usually with more distortion) and its reference (**Ref**) form. Computed local distortion energy within a molecule is derived from energy difference between the reference and target forms for each fragment. In addition, the local distortion energy can be determined and categorized by one of three approaches (**M1–M3**) with different fragmentation and coordinate manipulation schemes, depending on the purpose and system's complexity. The main features of these three approaches are summarized in the main text (see below), while the detailed description and discussion are given in Electronic Supplementary Information (ESI†).

Method 1 (M1) Fragmentation/Coordinate Scheme. Following many-body expansion approximation prevalently used in many fragmentation schemes,^{60–65} the total energy difference (E_{Diff}) between the target (**Tar**) and reference (**Ref**) molecules can similarly be estimated using fragmentation energy difference as in eq. 1.

$$E_{\text{Diff}} \approx \sum_i (E_i^{\text{Tar}} - E_i^{\text{Ref}}) + \sum_{i,j} (E_{ij}^{\text{Tar}} - E_{ij}^{\text{Ref}}) \dots \quad (1)$$

Where E_i and E_{ij} represent the one-body and two-body energy contribution, respectively. The distortion energy ($E_{\text{Distort},i}$) of each small fragment (i.e., local distortion) can be evaluated by considering only the one-body contribution after fragmentation of the two forms (**Tar** and **Ref**, Fig. 1b and Scheme S1†) in this **M1** method. Hence, local distortion energy is approximated by the distortion energy difference between the two forms of each fragment (eq. 2). In addition, distortion of all coordinates (e.g., bonds, angles and/or dihedrals) within each fragment are computed collectively in this **M1** method.

$$E_{\text{Distort},i} = E_i^{\text{Tar}} - E_i^{\text{ref}} \quad (2)$$

The size of fragment(s) in this **M1** method can be varied from the smallest possible fragment(s) (e.g., one heavy atom and its link atom(s)) to larger fragment(s) (including conjugate groups, e.g. aromatic rings or alkenyl/alkynyl) customized by the users. The minimum possible number of heavy atoms included in each fragment is often adopted to gain higher “resolution” of the distortion map and reduce the computational cost, if the key bonding interactions can be captured within the fragment. Moreover, such fragmentation processes also introduce link atoms (LAs) to cap all dangling bonds. It should be noted that local distortion energy is partially estimated by LA approximation, when its distorted boundary contains a LA.

Method 2 (M2) Fragmentation/Coordinate Scheme. Analogous to molecular mechanics, alternative fragmentation and decomposition of the distortion energy into three bonding terms (E_{bond} , E_{angle} and, optionally, E_{dihedral}) relating to each internal coordinate can be summarized in eq. 3. To achieve such distortion decomposition, fragmentation including two, three or four heavy atoms (with their link atom(s)) for one target internal coordinate (bond, angle or dihedral, respectively) taken from the reference form is performed to generate the **Ref** fragments. Then, only one target internal coordinate of each

ARTICLE

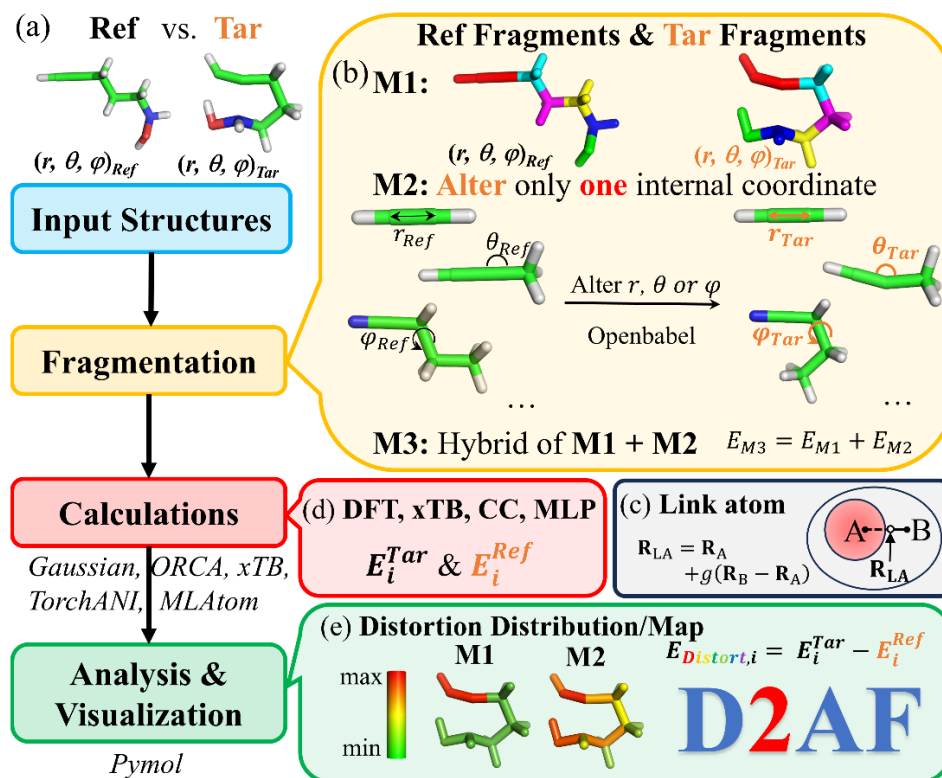


Fig. 1 (a) Our workflow of distortion distribution analysis with fragmentation (D2AF). The input structures contain the target system (**Tar**) and its reference (**Ref**) form. (b) Schematic fragmentation and coordinate manipulation methods (**M1-M3**). (c) ONIOM-type link atom (LA) treatment to "project" distortion-energy of a given covalent bond into the fragments. (d) Energy calculations of fragments. (e) Distortion distribution analysis and visualization.

generated fragment is altered to be identical to that coordinate value in the target system to set up all combinations of the **Tar** fragments (Fig. 1b and Scheme S1†). Accordingly, in contrast with the **M1** method, local distortion of each bonding coordinate within one fragment is evaluated individually.

$$\Delta E_{Distort} \approx \Sigma \Delta E_{bond} + \Sigma \Delta E_{angle} + \Sigma \Delta E_{dihedral} \quad (3)$$

Method 3 (M3) Fragmentation/Coordinate Scheme. Since some improper fragmentation in **M2** can neglect some important electronic effects (such as delocalization or lone pairs repulsion). Therefore, when dealing with such complex conjugated systems, a hybrid partition (**M3**) is applied for two different situations: any delocalized conjugated moiety using **M1** and the rest of the localized part using **M2** (Scheme S1†). Moreover, for challenging metal coordination systems, their complex metal-ligand interactions are hardly decomposed. This new hybrid features in **M3** allow the users to define a minimum metal-ligand coordination region as one special fragment to capture the key metal-ligand interactions using **M2**, while the remaining parts can be evaluated using **M1** and/or **M2** (Scheme S2†).

Link-atom Treatment. Fragmentation processes generally need to introduce link atoms (LAs) to cap all dangling bonds in all generated fragments, when a boundary involving single, double or triple bond is broken. The ONIOM-type boundary and LA approach are adopted.⁶⁶⁻⁶⁷ Therefore, the position of the link atom is not fixed, but depends on the bond distance of the two boundary atoms (Fig. 1c). In **M1**, such LA approach can project distortion of the bonding boundary to the link atom, whereas local distortion contribution should not rely on the LA approximation in **M2**.

Calculations. Energies of all generated fragments can then be computed using user-specified Quantum-Mechanics (QM) or Machine-Learning potentials (MLPs) method(s) and programs (Fig. 1d). Our current Python package interfaces a few QM and MLP programs (e.g., Gaussian,⁶⁸ ORCA,⁶⁹ xTB,⁷⁰ TorchANI,⁷¹ and MLatom⁷²).

Analysis and Visualization. After the energy calculations of each fragment, the distortion energy for the individual target and reference fragments can be determined on the basis of the **M1-M3** schemes. All distortion energies are recorded in an Excel file. Additionally, PyMOL scripts are generated to visualize the

ARTICLE

distribution of (relative) distortion energy (distortion map, Fig. 1e). Moreover, fragments in **M1** are colored according to their corresponding local distortion energies. Whereas, the local distortion energies in **M2** can be decomposed into specific bonding terms, and all bonds are color-coded based on the cumulative distortion energy contribution from all internal coordinates.

Computational Details. All these calculations were performed using our open-source python package (D2AF, <https://github.com/oscarchung-lab/D2AF>), which employs Open Babel package to treat internal coordinates of fragments.⁷³ The choice of the DFT method and basis sets as well as structures for each application system in this study were taken from the original studies (unless stated otherwise). All DFT calculations were conducted by Gaussian 16D1.⁶⁸ In addition, calculations using other computational methods for a few systems were also carried out: GFN2-xTB as semi-empirical method (SE) by xtb-python,⁷⁰ AIQM1 by MLatom 3.0,⁷⁴ and ANI-series by TorchANI⁷⁵⁻⁷⁶ as MLPs, as well as CCSD(T) by ORCA 5.0.^{69,77} (Relative) Distortion distribution (distortion map) were visualized by using PyMOL.⁷⁸

Results and discussion

Benchmark on Link-atom Approximation. In our study, fragmentation combined with ONIOM-type link atom (LA) treatment was adopted to estimate local distortion energy of a molecule (so-called distortion distribution). First, the effects of the LA approximation on the local distortion energy in **M1** were

examined. For distortion involving a single bond (e.g. C-C), a fragment with the popular hydrogen link atom (H-LA) can give similar energy profiles as the original ethane system with varying its bond length and angle (Fig. S1-S2†). In addition, for the case of distortion involving a C-C double-bond boundary in **M2**, uncommon carbon or oxygen link atom (C- or O-LA) was applied to ethene for evaluating these effects of double-bond link atom approximation (Fig. 2a-2c and S8-S14†). Compared to the computed energy profiles for the original ethene system, our M06-2X results show that C-LA exhibits a smaller energy deviation than O-LA for the C-H bond change (e.g., ~0.5 kcal/mol (C-LA) vs. ~10.1 kcal/mol (O-LA) at 1.4 Å; Fig. 2b). Regarding the H-C-H angle bending, O-LA and C-LA have an error of ~10 kcal/mol at large angle of 140° (Fig. 2c), whereas the former LA has a smaller error than the latter one for small H-C-H bond angles.

Moreover, similar evaluations were conducted for the uncommon link atom (N-LA or radical-type C-LA) associated with a C-C triple bond boundary of 1-butyne (Fig. 2d-2f). Our findings indicate that N-LA and C-LA give very similar energy curves for the distortion of altering C-C distance and C-C-C angle. Several benchmarks for LA approximation involving the other typical bonds encountered in this study were also investigated and these results show that LA approximation can qualitatively give reasonable distortion energy profiles (see Fig. S1-S16† for details). Based on our analysis of these benchmarks, H-LA, C-LA and N-LA were generally employed for the boundary involving single, double and triple bonds, respectively, unless stated otherwise.

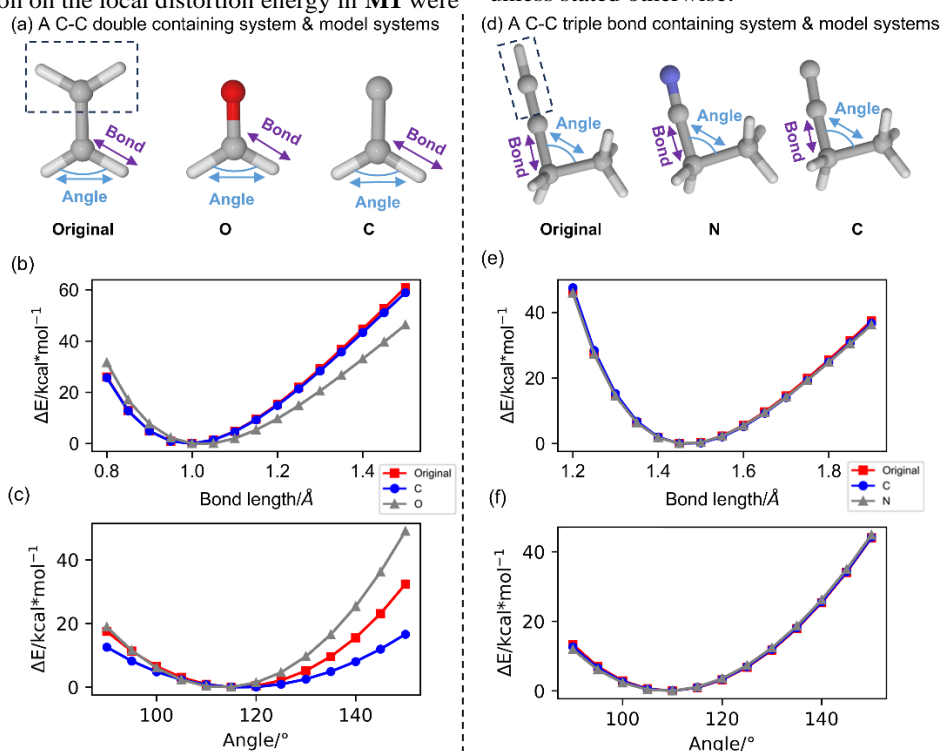


Fig. 2 Benchmark of link-atom approximation for C-C double (a-c) and triple (d-f) bonds. (a) Ethene as the original system and its two corresponding models (CH_2O , CH_2C) with either O or C as the link atom, as well as (d) 1-butyne as the original system and its two corresponding models ($\text{C}_3\text{H}_5\text{N}$, $\text{C}_3\text{H}_5\text{C}$ radical) with either N or radical-type C as the link atom, respectively. Energy profiles for their varying bond length and angle derived from rigid scan calculations of the real systems followed by replacement by the link atom and single-point energy calculations computed at M06-2X/6-31G(d) level.

ARTICLE

Applications of Distortion Distribution in Different Systems

In order to demonstrate the robustness and broad applicability of our approach, a diverse set of chemical and biological systems were systematically investigated. The key results for the nine representative systems will be presented in the main text, while their detailed results as well as the other 11 systems are provided in ESI†. The representative systems can be categorized into different types: (i) organic systems (reverse Cope elimination,⁷⁹ Diels-Alder cycloaddition,⁸⁰ azide-alkyne cycloaddition⁸¹ as well as triplet excited-state Di- π -methane rearrangement reactions);⁸²⁻⁸³ (ii) the key part of a supramolecular system (an addition reaction within a macrocycle[2]rotaxane);⁸⁴ (iii) metal-coordination system (Ir-catalyzed C-H borylation);⁸⁵ (iv) biological systems (an artificial heme enzyme catalyzed Si-H insertion,⁸⁶ a Spleen tyrosine kinase protein in complex with an Imatinib drug). In addition, multiple structures obtained from MD simulations and IRC calculations of two simple S_N2 reactions were also applied to show the change in (relative) distortion distribution during the reaction process.

Reverse Cope Elimination Reaction (Organic System). Fig. 3a presents an intramolecular concerted addition reaction, which involves the addition of the proton and the amine onto the triple bond along with the O-H bond breaking. The Houk group systematically studied this reaction mechanism.⁷⁹ Their DFT results showed that the main distortion comes from the alkyne (~17.6 kcal/mol) and hydroxylamine (~14.5 kcal/mol) parts along with some distortion of the tether part (~8.9 kcal/mol). Comparatively, our **M1** and **M2** schemes are also applied to get detailed insight of distortion distribution of this reaction (Fig. 3b-3g). Distortion distribution using **M1** also supports that the most significant distortion originates from the alkyne moiety (~17.6 kcal/mol, Fig. 3c). Distortion distribution by **M2** generally shows a similar distribution (Fig. 3e), i.e., bending of the triple bond as the main distortion contributor ($\Delta(\mathbf{A1})$: -41.9° with distortion energy of ~9.6 kcal/mol; $\Delta(\mathbf{A2})$: -19.6° with distortion energy of ~3.6 kcal/mol).

Similar to the previous Houk study, distortion of the hydroxylamine part is also significant in the **M2** scheme, due to the O-H bond stretching ($\Delta(\mathbf{B1})$: 0.12 Å with distortion energy of ~6.3 kcal/mol) and the O-N bond compression ($\Delta(\mathbf{B2})$: -0.09 Å with distortion energy of ~2.8 kcal/mol). In contrast, a smaller distortion energy of the hydroxylamine part (~1.5 and ~1.6 kcal/mol, Fig. S18†) was shown in the **M1** results than our benchmark for this O-N bond boundary (Fig. S3†), when the (default) smallest fragmentation strategy was used (splitting the OH and NH parts). These results indicate that the H-link atom underestimates destabilization of the O/N lone pairs repulsion within the hydroxylamine part. Pleasingly, slight expansion of that fragment by including the OH and NH parts leads to more

reliable results (Fig. 3c and S12). Moreover, a considerable dihedral change of the tether ($\Delta(\mathbf{D1})$: -170.8°) motivated us to include its distortion energy contribution (~4.9 kcal/mol) in the **M2** scheme and expand one methylene fragment in the **M1** scheme. Such unfavourable eclipsed conformation relating to **D1** in **S1-TS** is in agreement with the findings of the Houk group. Consequently, our distortion analysis using the **M1** and **M2** approaches can reveal a detailed and complementary picture of distortion distribution, which can gain more mechanistic understanding.

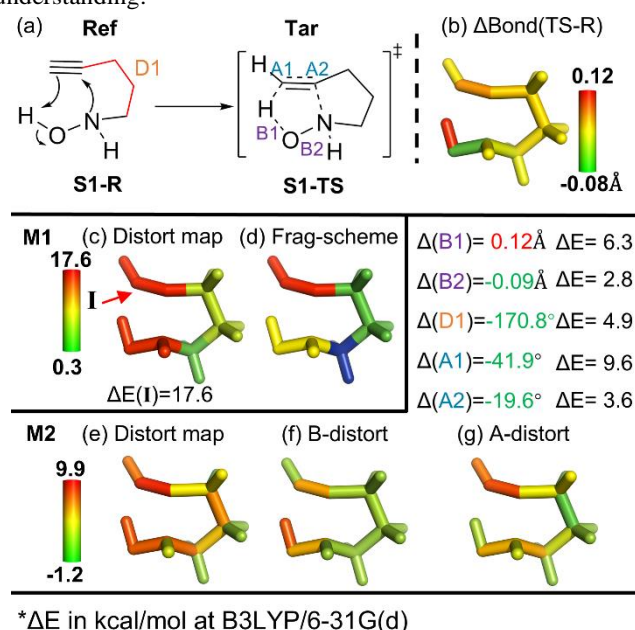


Fig. 3 (a) Overview of a reverse Cope elimination (structures taken from ref 79). (b) Bond length change (Δ Bond in Å) from the reference form (**Ref**; reactant, **S1-R**) to the target form (**Tar**; TS, **S1-TS**). (c) Distortion distribution (kcal/mol) and (d) fragmentation using **M1** scheme. (e) Total distortion distribution (kcal/mol), its distortion contribution from (f) bond and (g) angle terms using **M2** scheme. The major local distortion contributed from the dihedral (**D1**) was also considered by using an expanded CH₂CH₂ fragment in **M1** scheme and adding **D1** in **M2** scheme). The O-N group was also considered as one fragment. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

Diels-Alder Reaction (Organic System). Our second system is a Diels-Alder reaction between cyclopentenone and cyclopentadiene.⁸⁰ As illustrated in Fig. 4c and 4e, our analysis results suggest that the major distortion is primarily contributed by the addition sites. In addition, our benchmark shows that the distortion energy contributed by the small C-C-C angle **A1** cannot be well reproduced by H-LA (Fig. S4†). Accordingly, the fragment **I** containing the **A1** angle results in modest underestimation of its distortion energy (~1.3 kcal/mol) using **M1** scheme with H-LA. That distortion energy of fragment **I** can raise to ~3.0 kcal/mol (Fig. 4c), when a larger CH group was used as the link atom (Fig. S4†). Similarly, the fragment related to the small **A1** angle also shows a reasonable energy (~3.1

ARTICLE

kcal/mol) in the **M2** scheme, only when keeping the CH-CH₂-CH group without LAs is adopted (Fig. S15†).

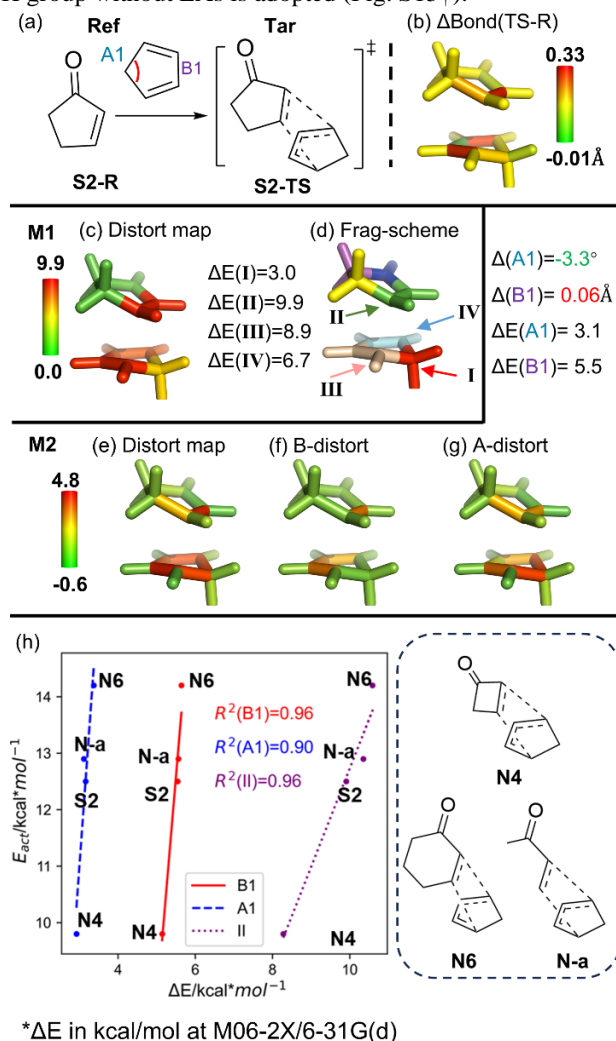


Fig. 4 (a) Overview of a Diels-Alder cycloaddition (structures taken from ref 80). (b) Bond length change (Δbond in Å) from the reference form (Ref; reactant, **S2-R**) to the target form (Tar; TS, **S2-TS**). (c) Distortion distribution (kcal/mol) and (d) fragmentation using **M1** scheme. (e) Total distortion energy distribution (kcal/mol), its distortion contribution from (f) bond and (g) angle terms using **M2** scheme. The CH was chosen as the link atom for the fragment **I** in **M1** and no link atom was applied to **A1** angle in **M2**. (h) Correlation between the computed local distortion energy of the key fragments (**A1**, **B1**, **II**) and activation energy of the Diels-Alder reaction with four different substrates. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

Besides, the Houk group pioneeringly discovered a strong linear correlation between the molecular distortion and activation energy in a series of [4+2] addition reactions ($R^2 = 0.93$).⁸⁰ To further demonstrate the efficacy and usefulness of our analysis, good linear correlations between the activation energy and one of the main local distortion components (fragment **II** in **M1** scheme; **B1** or **A1** in **M2** scheme) are also observed (Fig. 4h, $R^2 = 0.90$ - 0.96). These results indicate that the main local distortion not only helps understand the main source of the barrier, but also can potentially serve as a new descriptor/feature

for multivariate linear regression (MLR) and machine learning (ML) modeling.⁵⁹

Azide-alkyne Cycloaddition (Organic System). Our distortion distribution was also carried out to analyze the strain-promoted [3+2] dipolar cycloaddition between azide and strained 2-methyloxyl-cyclooctyne,⁸¹ which is a vital bio-orthogonal reaction.⁸⁷ As shown in Fig. 5b, the most significant distortion using the **M1** scheme was found to be the azide (~ 15.3 kcal/mol) and alkyne (~ 4.9 kcal/mol) moieties. On the other hand, analysis using **M2** is not easy to manipulate only one internal coordinate within the cyclic moiety. Instead of **M2**, the hybrid **M3** scheme was thus employed in this cyclic system (Fig. 5e-5g). Likewise, our **M3** results further support these two major distortion contributors, due to considerable bending of the azide ($\Delta(\text{A1})$: -31.5° , $\Delta E(\text{A1})$: ~ 13.8 kcal/mol) and N-N bond stretching ($\Delta(\text{B1})$: 0.02Å , $\Delta E(\text{B1})$: ~ 4.1 kcal/mol). In addition, bending of the reacting alkyne regions (**A2** and **A3**) contributes to some distortion (~ 2.4 and 3.3 kcal/mol, respectively).

To get deeper understanding on this strain-promoted addition, the same reaction with an acyclic alkyne substrate (2-butyne, **S3*-R**) has also been studied by the same computational method in this work (Fig. 5h). A higher barrier (~ 20.6 kcal/mol) for this acyclic alkyne was found in our calculations, compared to the strained alkyne (~ 7.7 kcal/mol). Notably, our D2AF results highlight that distortion energies of **A1** (~ 18.9 kcal/mol) and **A2** (~ 4.5 kcal/mol) in **S3*-TS** are increased relative to the **S3*-R** substrate, compared to those for **S3-TS** and **S3** ($\Delta E(\text{A1})$: ~ 14.6 kcal/mol and $\Delta E(\text{A2})$: ~ 2.8 kcal/mol). To further compare with the strained alkyne (**S3-R**), **S3_{model}-R** and **S3_{model}-TS** models were generated by truncating **S3-R** and **S3-TS** (only keeping the alkyne part) followed by capping with H-LA. Single-point energies on the truncated **S3_{model}-R** and **S3_{model}-TS** models give a similar barrier (~ 9.6 kcal/mol), which is comparable to the full **S3** system (~ 7.7 kcal/mol). Moreover, **S3_{model}-R** and **S3_{model}-TS** models were found to have higher distortion energies than **S3*-R** and **S3*-TS** by roughly ~ 16.3 and ~ 5.3 kcal/mol, respectively. Accordingly, much higher distortion energy in the cycloalkyne **S3-R** reactant (i.e. reactant destabilization or pre-organization effect) should be the key factor in lowering the barrier. In addition, a larger reaction driving force (ΔE_{RXN} , partly driven by strain relief⁸⁸ for **S3_{model}** than **S3*** (-90.7 and -80.4 kcal/mol, respectively) should be another factor in lowering the barrier and rendering earlier transition state (with a smaller distortion energy based on Hammond postulate).

Triplet Excited-state Di- π -methane Rearrangement (Organic System). Apart from the previous reaction systems in a closed-shell singlet ground state, a triplet excited-state Di- π -methane rearrangement of benzobarrelene involving spins was selected for our distortion analysis (Fig. 6a).⁸² Our **M1** scheme reveals that the key distortion should result from the bridgehead fragment **I** (Fig. 6c-6d, $\Delta E(\text{I})$: ~ 8.7 kcal/mol). Likewise, the **M3** scheme shows that the C-C-C bending between the two reacting

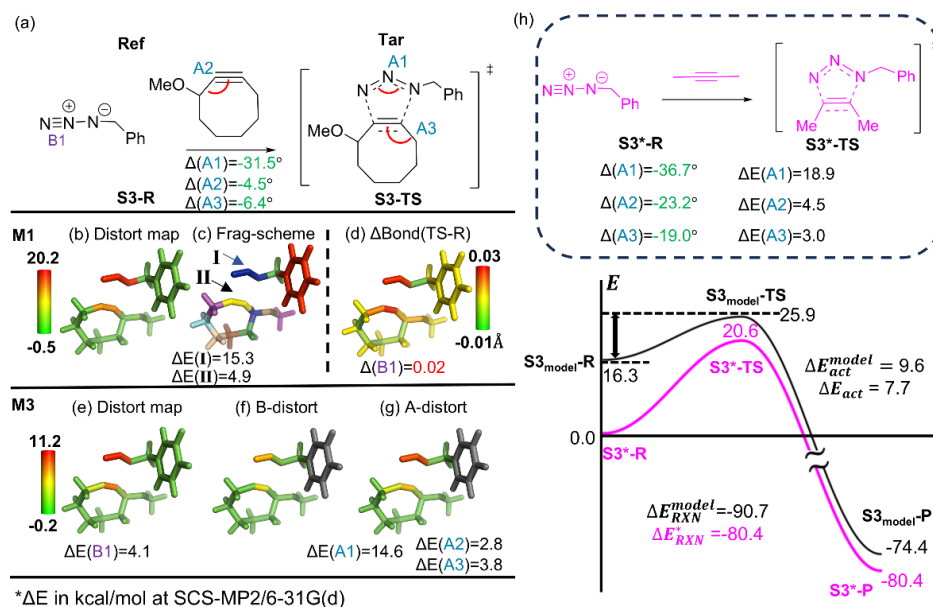


Fig. 5 (a) Overview of the azide-alkyne cycloaddition (structures taken from ref 81). (b) Distortion distribution (kcal/mol) and (c) fragmentation using **M1** scheme. (d) Bond length change (Δbond in Å) from the reference form (**Ref**; reactant, **S3-R**) to the target form (**Tar**; TS, **S3-TS**). (e) Total distortion distribution (kcal/mol) (f) distortion contribution from bond and (g) angle using **M3** scheme. (h) Analysis using the truncated **S3_{model}** (extracting from **S3-R** and **S3-TS** with keeping only the alkyne part) to compare with the energy for the 2-butyne (**S3***) substrate. The key geometrical changes and their corresponding distortion energy (ΔE) are also given. ΔE_{act} refers to the activation energy for the original **S3** substrate.

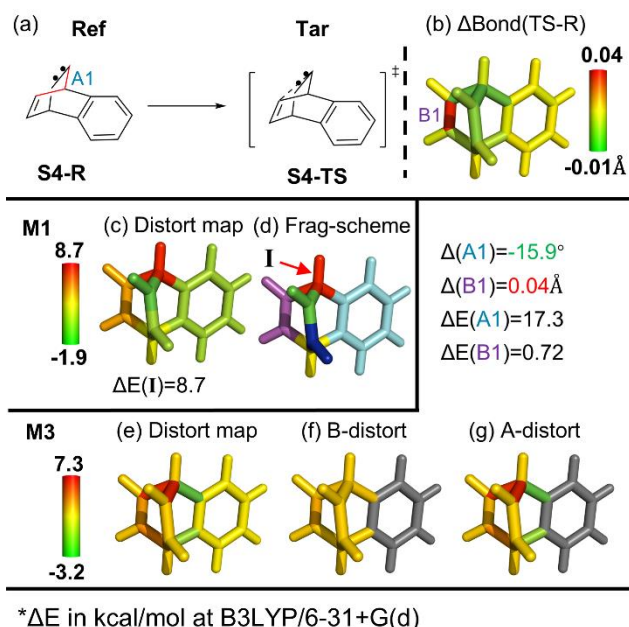


Fig. 6 (a) Overview of the intermediate in a triplet Di- π -methane intramolecular rearrangement (structures taken from ref 82). (b) Bond length change (Δbond in Å) from the reference form (**Ref**; reactant, **S4-R**) to the target form (**Tar**; TS, **S4-TS**). (c) Distortion distribution (kcal/mol) and (d) fragmentation using **M1** scheme. (e) Total distortion distribution (kcal/mol) (f) distortion contribution from (f) bond and (g) angle using **M3** scheme. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

regions ($\Delta(A1)$: -15.9° , $\Delta E(A1)$: ~ 17.3 kcal/mol) results in the most significant distortion within the molecule (Fig. 6e-6g). While stretching of the reacting C=C bond ($\Delta(B1)$: 0.04 Å) contributes to the trivial distortion (~ 0.7 kcal/mol). Therefore, our analysis points out that the bending angle (**A1**) plays a pivotal role, which may not straightforwardly be seen. Additionally, these findings highlight a certain capability of our analysis to excited-state systems.

Cyclization of Fumaramide within [2]Rotaxane (Supramolecular System). Recently, supramolecular catalysis has attracted more attention.⁸⁹ We extracted the two key intermediate (guest) molecules from our recent study on stereoselective cyclization of fumaramide into *trans*- β -lactam within [2]Rotaxane (Fig. 7a).⁸⁴ Our previous study showed that the stability of the two key intermediates (**S5-cis** and **S5-trans**) determined the stereoselectivity, in which the *cis* intermediate has a larger distortion energy than the *trans* intermediate by ~ 5.0 kcal/mol. Our distortion distribution of the *cis* (minor) intermediate is further compared to the *trans* (major) intermediate (Fig. 7a). In contrast to the abovementioned organic systems, our initial analysis using **M1** or **M3** (considering contribution from bonds and angles only) scheme shows that the sum of the local distortion energy components ($< \sim 1.0$ kcal/mol) are obviously smaller than the global distortion energy of the whole intermediates (~ 5.0 kcal/mol, Fig. S25 and Table S1†).

ARTICLE

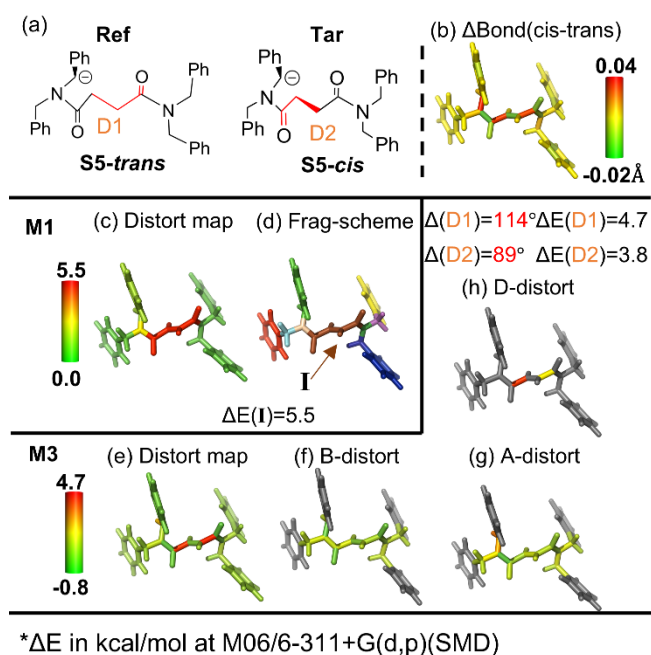


Fig. 7 (a) Overview of the *cis* and *trans* isomers of an intermediate during the CsOH-promoted cyclization (structures taken from ref 84). (b) Bond length change (Δbond in Å) from the reference form (**Ref**; **S5-trans**) to the target form (**Tar**; **S5-cis**). (c) Distortion distribution (kcal/mol) and (d) fragmentation using **M1** scheme. (e) Total distortion distribution (kcal/mol), distortion contribution from (f) bond, (g) angle and (h) dihedral, using **M3** scheme. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

These results imply that dihedral(s) can be the major distortion contributor, due to the large change in the **D1** and **D2** dihedrals (Fig. 7a-7c, (Δ(**D1**) = 89° and Δ(**D2**) = 114°). Therefore, a larger fragment **I** (including **D1** and **D2** dihedrals) was further treated using the **M1** scheme (Fig. 7d). A higher distortion energy of this key fragment (ΔE(**I**) = ~7.8 kcal/mol) is primarily attributed to tremendous differences in the two dihedrals between these two intermediates (Fig. 7a-7c). Similarly, upon incorporation of the contribution from these dihedral components using the **M3** scheme, these two key dihedrals (**D1** and **D2**) leading to distortion energy of ~6.0 and ~4.7 kcal/mol are identified (Fig. 7e-7f), respectively. This analysis confirmed that the *trans*-stereoselective cyclization is attributed to the higher distortion mainly caused by the two key dihedrals of the *cis* form enforced by the rotaxane.

Ir-catalyzed C-H Borylation (Coordination System). Besides the above organic systems, we also applied local distortion analysis on a metal-coordination system, Ir-catalyzed C-H activation consisting of (bpy)Ir(Beg)₃ (bpy: 2,2'-bipyridine; Beg: (ethyleneglycolato)boron) and indole (Fig. 8a).⁸⁵ It should be noted that fragmentation on metal-coordination systems is scarce and much less straightforward. To maintain complex metal-ligand bonding around the metal center, the Ir and three ligated boron atoms as well as the bpy ligand were set as one specialized fragment (see Scheme S2 and Fig. S28† for the detailed fragments and coordination region). A similar fragmentation strategy to keep the metal-ligand environment/interactions for the other coordination systems were also utilized.

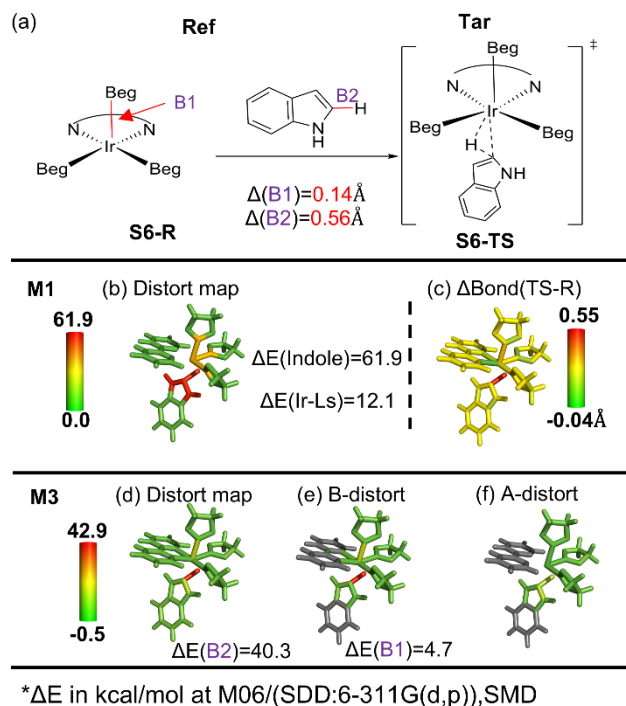


Fig. 8 (a) Overview of iridium catalyzed C-H borylation (structures from ref 85). (b) Distortion distribution (kcal/mol) using **M1** scheme and (c) bond length change (Δbond in Å) from the reference form (**Ref**; reactant, **S6-R**) to the target form (**Tar**; TS, **S6-TS**). (d) Total distortion distribution (kcal/mol), (e) distortion contribution from bond and (f) angle using **M3** scheme. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

Our **M1** analysis reveals that the Ir coordination region and indole mainly contribute the distortion energy (~12.1 and ~61.9 kcal/mol, respectively; Fig. 8b-8c). What's more, a "finer" distortion map can be attained by using the **M3** scheme (Fig. 8e and 8g). One of the three Ir-B bonds (**B1** in Fig. 8a) exhibits the longest Ir-B bond elongation by 0.14 Å, resulting in a distortion energy of ~4.7 kcal/mol. On the other hand, the reacting C-H bond of the indole ring experiences a very high distortion energy of ~40.3 kcal/mol, due to its bond elongation by 0.55 Å (Fig. 8a and 8e). Consequently, this analysis using both **M1** and **M3** schemes can offer different and complementary "resolution" on the distortion distribution. Moreover, our findings on local distortion energy for the key Ir and indole moieties are qualitatively consistent with the previous study, in which computed distortion energy of the entire Ir catalyst and indole molecule were reported to be about 9.8 and 63.9 kcal/mol by the Houk group, respectively.⁸⁵

Si-H Insertion in Artificial Heme Enzyme (Biological System). Directed-evolution artificial heme enzymes pioneered by the Arnold group play a critical role in new biocatalysis.⁹⁰ They recently developed an artificial heme protein to catalyze novel Si-H insertion (Fig. 9a), which is selected as an example of our biological systems.⁸⁶ To reduce computational cost and get a higher "resolution" distortion map, a simplified heme region (red fragment **II** in Fig. 9c; see Fig. S29† for the detailed fragments) is utilized as one fragment which was previously proposed to represent the heme.⁹¹ Our **M1** analysis results unveil that the reacting Si-H part (fragment **I**) and the heme coordination region (fragment **II**) mainly contribute to distortion energy (~6.3 and

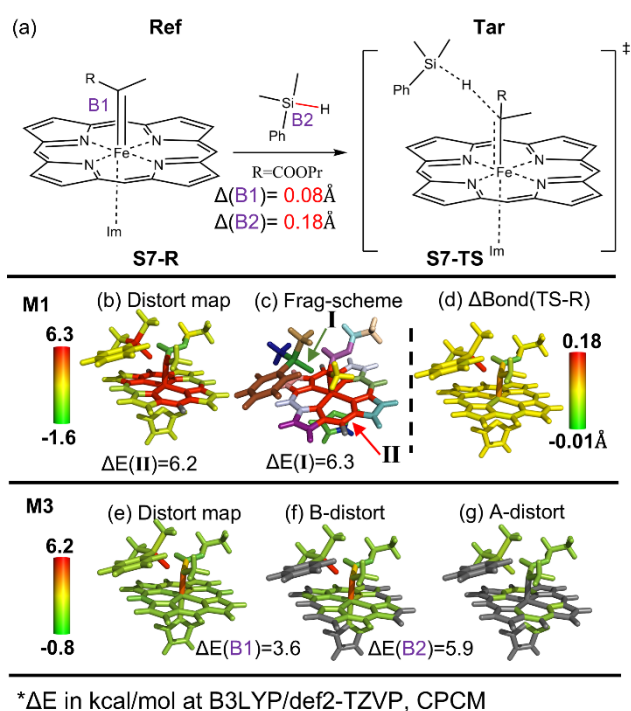


Fig. 9 (a) Overview of the heme protein catalyzed Si-H insertion (structure from ref 86). (b) Distortion distribution (kcal/mol) and (c) fragmentation using **M1** scheme. (d) Bond length change (ΔBond in \AA) from the reference form (**Ref**; reactant, **S7-R**) to the target form (**Tar**; TS, **S7-TS**). (e) Total distortion distribution (kcal/mol), distortion contribution from (f) bond and (g) angle using **M3** scheme. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

~6.2 kcal/mol, respectively, Fig. 9b-9c). Furthermore, as shown in Fig. 9d, our **M3** results illustrate a comparable distortion energy of ~3.6 and ~5.9 kcal/mol for the elongation of the Fe-C

bond ($\Delta(B1)$: 0.08 \AA) and Si-H bond ($\Delta(B2)$: 0.18 \AA), respectively. Consequently, **M1** exhibits distortion of the larger coordination region, while **M3** further “localizes” the main distortion into the Fe-C and Si-H bonds.

Protein-drug Binding (Biological System). In order to further illustrate structural improvement achieved through multi-scale quantum refinement (QR), the refined Imatinib drug structure inside the Spleen tyrosine kinase protein structure after our recent quantum refinement⁹² is taken as the reference state, while its distorted X-ray (XR) structure serves as the target state (Fig. 10). The drug structure improved by QR was found to exhibit a significantly lower distortion energy than its XR structure (~22.6 kcal/mol).⁹² Our **M1** analysis unravels that the main distortion of the XR structure originates from the pyrimidine (Ring1) and pyridine (Ring2) rings (ΔE : ~9.8 and ~3.1 kcal/mol, respectively; Fig. 10c). Besides, our **M3** results further delineate a distortion energy of ~2.7 kcal/mol derived from changing the bond distance (**B1**) between the pyridine and pyrimidine ring by -0.07 \AA . Interestingly, its neighbouring Met448 and Gln449 residues are found to have very close contacts with the pyridine and pyrimidine rings of the drug. These contacts become longer ($\Delta(r1)$: 0.02 \AA and $\Delta(r2)$: 0.18 \AA) and the drug releases its distortion energy simultaneously after QR. Therefore, these two residues may enforce less distortion influence on these two rings. Additionally, as some subtle difference in the local distortion for the amide group between **M1** and the default **M3** results was also found (Fig. S31†), **D1** was further added in the **M3** scheme and gave a distortion energy of ~4.9 kcal/mol ($\Delta(D1) = 36.6^\circ$). Such analysis further demonstrates more insights into biological systems gained by our local distortion analysis and supports the performance of QR in improving the local protein-drug structure.

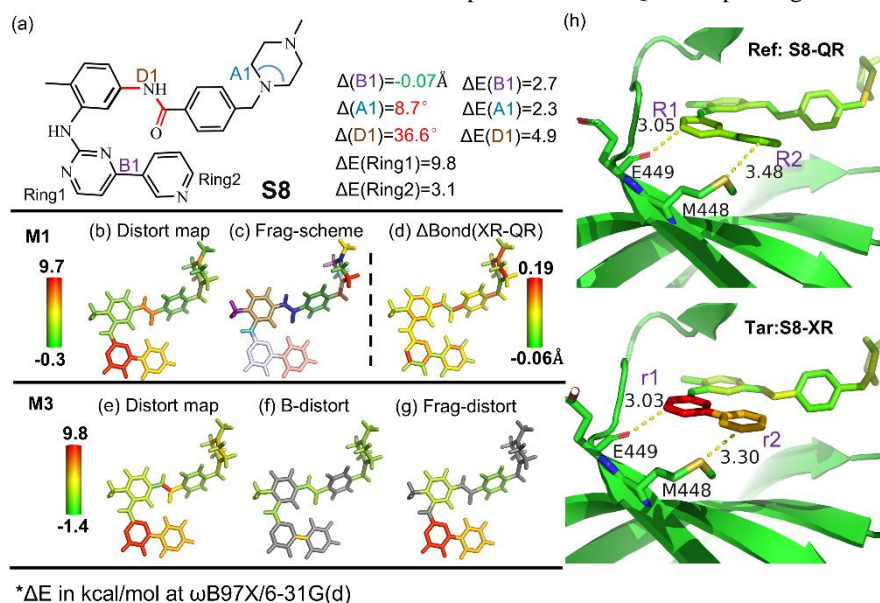


Fig. 10 (a) Overview of structures of Imatinib-Spleen tyrosine kinase (structures from ref 92). (b) Distortion distribution (kcal/mol) and (c) fragmentation using **M1** scheme. (d) Bond length change (ΔBond in \AA) from quantum refinement (**Ref**, QR, **S8-QR**) structure to X-ray (**Tar**, XR, **S8-XR**) structure. (e) Total distortion distribution (kcal/mol), distortion contribution from (f) bond and (g) angle using **M3** scheme. (h) Scheme for binding site interaction with neighbouring residues taken from the XR and QR structures. The drug structures are colored based on their relative distortion to that optimized structure in gas phase using **M1** scheme. The key geometrical changes and their corresponding distortion energy (ΔE) are also given.

ARTICLE

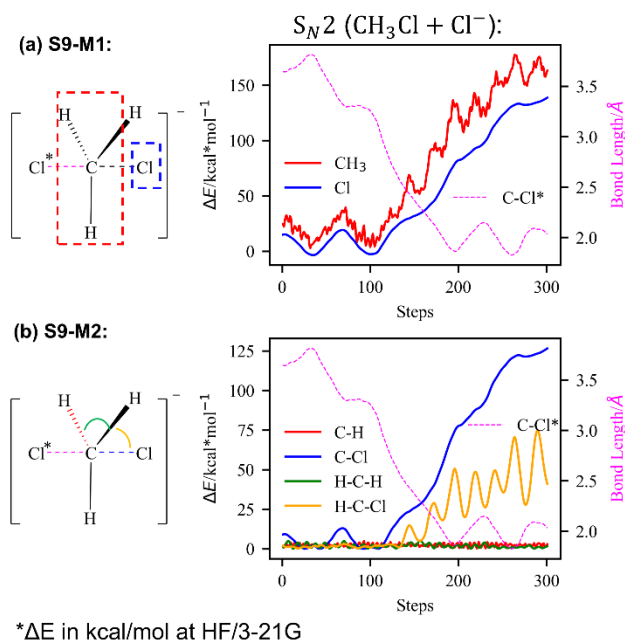


Fig. 11 Time-evolved distortion distribution (kcal/mol) of a MD trajectory for S_N2 reaction between CH_3Cl and a Cl^- (denoted as Cl^*) using (a) **M1** scheme and (b) **M2** scheme at HF/3-21G level. Each trajectory averages distortion energy belonging to the same kind of chemical bonds/angles. C- Cl^* curve (purple dotted line) indicates the distance between Cl^* and center C.

Molecular Dynamics of S_N2 Reactions (Multiple Structures). To further validate the flexibility and generalization of our distortion analysis approach, multiple structures taken from molecular dynamics (MD) simulations (Fig. 11) and IRC process (Fig. S67†) of two simple S_N2 reactions were chosen as the target states to analyze (relative) distortion energy change throughout the reaction process. During the new C- Cl^* bond forms in the MD simulation (purple curve in Fig. 11), our **M1** analysis shows that the CH_3 fragment and the leaving Cl fragment exhibit increasing trends of the distortion energy as Cl^* anion approaches (Fig. 11a). Our **M2** analysis further decouples the distortion energy and shows that the principal distortion energies are contributed from the C-Cl bond stretching and the H-C-Cl bending (the yellow and blue curves in Fig. 11b) during the C-Cl bond rupture. A similar distortion picture is also found in the IRC calculations on a similar S_N2 reaction (Fig. S68-S69†). Overall, this additional analysis shows the flexibility and generalization of our method to reaction processes involving multiple structures.

Different Computational Methods. Finally, our flexible local distortion analysis can also be applied to different computational chemistry methods and/or software packages to estimate fragments' energies. A few methods (highly-accurate CCSD(T), cost-effective semi-empirical GFN2-xTB or machine learning potentials (MLPs, e.g., ANI-2x) methods)^{70,76-77} were employed to evaluate the distortion maps of our four representative systems

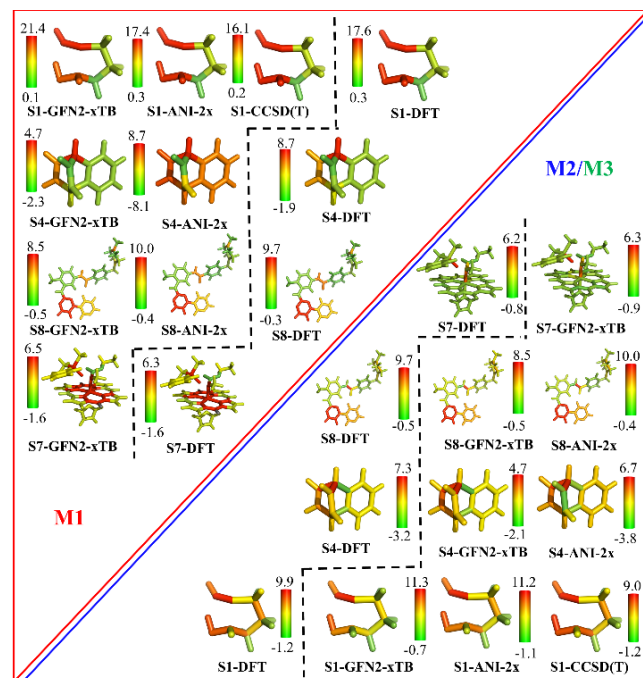


Fig. 12 Distortion distribution (kcal/mol) of the four selected systems calculated by DFT, SE (GFN2-xTB), MLP (ANI-2x) and CCSD(T) (only for **S1**) methods

(Fig. 12). To the best of our knowledge, this is the first study using CCSD(T) or MLP to evaluate distortion energy. Generally, GFN2-xTB shows consistent distortion maps with the DFT- or CCSD(T)-computed results. The ANI-2x method performs well except for the challenging triplet-state system (**S4**), due to no spin systems used for ANI-2x training. Therefore, the efficient GFN2-xTB method is potentially used for a wider range of systems, while MLPs could be applied to some neutral ground-state systems.

Conclusions

In this study, a general, efficient and flexible approach for local distortion distribution analysis based on fragmentation and ONIOM-type boundary strategies was proposed and implemented. Benchmarks for several single-, double- and triple-bond link-atom approximations were first performed to support reasonable boundary treatments. Distortion distribution of various chemical (including organic, supramolecular, inorganic and coordination systems) and biological (natural and artificial enzymes as well as proteins with drugs) systems were then extensively examined and compared. Our analysis can successfully reveal qualitative (relative) distortion distribution of the target structure(s) compared to the reference structure, in which any type of structures (e.g., local minimum, transition

state(s), MD- or IRC-derived structures) can be adopted and compared. Additionally, local distortion distribution on multiple structures or different computational methods (such as high-level CCSD(T), cost-efficient xTB and MLP methods) to compute the fragments' energy can readily be applied and compared to support our general, efficient and flexible analysis. Our analysis on diversified molecules shows potent feasibility to aid people in gaining deeper understanding and more insights into structures, reaction mechanisms and dynamics. Furthermore, the key local distortion energy can potentially be utilized as a new descriptor in multi-linear regression (MLR) or machine learning (ML) modelling.

Moreover, local distortion can be "mapped" into different fragments or chemical bonds by three fragmentation/coordinate **M1-M3** schemes, which gives us more intuitive viewpoint. **M1-M3** schemes have their advantages and disadvantages. It is recommended that users use the **M1** scheme to obtain a broader view of distortion distribution, as well as use **M2** (for systems without conjugated or coordination parts) or **M3** (those containing conjugated or coordination parts) scheme to obtain a more detailed/localized distortion distribution along each coordinate. Serious difference of distortion map derived by **M1** and **M2** (or **M3**) in some regions may indicate mis-counting local distortion energy. Users can modify the fragment(s) for these regions to conduct further analysis, minimize the discrepancy and gain more understanding.

Data availability

The source code, input and output files of all the test systems have been deposited in our GitHub repository (<https://github.com/oscarchung-lab/D2AF>). All cartesian coordinates and key absolute energy of the new optimized structures are provided in electronic supplementary information (ESI).

Author contributions

L.W.C. conceived and supervised the project. Z.Y. and Y.S.L. developed and tested the D2AF code. Y.S.L. performed the link atom benchmark. Y.S.L. and Z.Y. carried out the D2AF calculations and analysis of different systems. Z.Y. and Y.S.L. prepared the original draft and further reviewed and edited by L.W.C and L.X. All authors analyzed and discussed the results as well as assisted manuscript preparation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (21933003, 22193020 and 22193023), Southern University of Science and Technology, the Shenzhen Nobel Prize Scientists Laboratory

Project (C17783101), Guangdong Provincial Key Laboratory of Catalysis (2020B121201002), and Natural Science Foundation of Shenzhen Innovation Committee (JCYJ20220530115408019). We thank the Center for Computational Science and Engineering of Southern University of Science and Technology CHEM HPC at SUSTech for partly supporting this work.

Notes and references

1. K. J. Laidler and M. C. King, *J. Phys. Chem.*, 1983, **87**, 2657-2664.
2. R. A. Marcus, *J. Chem. Phys.*, 1956, **24**, 966-978.
3. K. Fukui, *Angew. Chem., Int. Ed.*, 1982, **21**, 801-809.
4. R. B. Woodward and R. Hoffmann, *Angew. Chem., Int. Ed.*, 1969, **8**, 781-853.
5. A. Pross and S. S. Shaik, *Acc. Chem. Res.*, 1983, **16**, 363-370.
6. F. Weinhold, *J. Comput. Chem.*, 2012, **33**, 2363-2379.
7. G. Knizia, *J. Chem. Theory Comput.*, 2013, **9**, 4834-4843.
8. J.-X. Zhang, F. K. Sheong and Z. Lin, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2020, **10**, e1469.
9. I. Fernandez and F. M. Bickelhaupt, *Chem. Soc. Rev.*, 2014, **43**, 4953-4967.
10. F. M. Bickelhaupt and K. N. Houk, *Angew. Chem., Int. Ed.*, 2017, **56**, 10070-10086.
11. D. H. Ess and K. Houk, *J. Am. Chem. Soc.*, 2007, **129**, 10646-10647.
12. S. Nagase and K. Morokuma, *J. Am. Chem. Soc.*, 1978, **100**, 1666-1672.
13. K. Fukui, *Acc. Chem. Res.*, 1981, **14**, 363-368.
14. R. Ramanan, K. D. Dubey, B. Wang, D. Mandal and S. Shaik, *J. Am. Chem. Soc.*, 2016, **138**, 6786-6797.
15. L. C. Burrows, L. T. Jesikiewicz, G. Lu, S. J. Geib, P. Liu and K. M. Brummond, *J. Am. Chem. Soc.*, 2017, **139**, 15022-15032.
16. K. Morokuma, *Acc. Chem. Res.*, 1977, **10**, 294-300.
17. B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887-1930.
18. E. D. Glendening, *J. Phys. Chem. A*, 2005, **109**, 11936-11940.
19. Y. Mo, P. Bao and J. Gao, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6760-6775.
20. L. Zhao, M. von Hopffgarten, D. M. Andrada and G. Frenking, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1345.
21. P. Su, Z. Tang and W. Wu, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2020, **10**, e1460.
22. Y. Mao, M. Loipersberger, P. R. Horn, A. Das, O. Demerdash, D. S. Levine, S. Prasad Veccham, T. Head-Gordon and M. Head-Gordon, *Annu. Rev. Phys. Chem.*, 2021, **72**, 641-666.
23. G. Bistoni, A. Altun, Z. Wang and F. Neese, *Acc. Chem. Res.*, 2024, **57**, 1411-1420.
24. E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498-6506.
25. V. R. Cooper, L. Kong and D. C. Langreth, *Phys. Procedia*, 2010, **3**, 1417-1430.
26. O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.*, 2010, **133**, 244103.
27. S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105-5154.
28. A. Altun, F. Neese and G. Bistoni, *J. Chem. Theory Comput.*, 2019, **15**, 5894-5907.

ARTICLE

29. W. Yang and R. G. Parr, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 6723-6726.
30. P. Geerlings, F. De Proft and W. Langenaeker, *Chem. Rev.*, 2003, **103**, 1793-1874.
31. R. F. Bader, *Acc. Chem. Res.*, 1985, **18**, 9-15.
32. R. Maji, S. C. Mallojjala and S. E. Wheeler, *Acc. Chem. Res.*, 2023, **56**, 1990-2000.
33. L. Falivene, Z. Cao, A. Petta, L. Serra, A. Poater, R. Oliva, V. Scarano and L. Cavallo, *Nat. Chem.*, 2019, **11**, 872-879.
34. K. C. Harper, E. N. Bess and M. S. Sigman, *Nat. Chem.*, 2012, **4**, 366-374.
35. A. V. Brethomé, S. P. Fletcher and R. S. Paton, *ACS Catal.*, 2019, **9**, 2313-2323.
36. M. J. Phipps, T. Fox, C. S. Tautermann and C. K. Skylaris, *Chem. Soc. Rev.*, 2015, **44**, 3177-3211.
37. K. B. Wiberg, *Angew. Chem., Int. Ed.*, 2003, **25**, 312-322.
38. J. P. Meyer, P. Adumeau, J. S. Lewis and B. M. Zeglis, *Bioconjugate Chem.*, 2016, **27**, 2791-2807.
39. E. Kayahara, R. Qu and S. Yamago, *Angew. Chem., Int. Ed.*, 2017, **56**, 10428-10432.
40. J. Lan, X. Li, Y. Yang, X. Zhang and L. W. Chung, *Acc. Chem. Res.*, 2022, **55**, 1109-1123.
41. X. Zhang, L. W. Chung and Y.-D. Wu, *Acc. Chem. Res.*, 2016, **49**, 1302-1310.
42. I. Fernández, F. M. Bickelhaupt and D. Svatunek, *J. Chem. Theory Comput.*, 2023, **19**, 7300-7306.
43. D. Svatunek and K. N. Houk, *J. Comput. Chem.*, 2019, **40**, 2509-2515.
44. L. Wang, Z. Zeng, W. Gao, T. Maxson, D. Raciti, M. Giroux, X. Pan, C. Wang and J. Greeley, *Science*, 2019, **363**, 870-874.
45. R. Gianatassio, J. M. Lopchuk, J. Wang, C.-M. Pan, L. R. Malins, L. Prieto, T. A. Brandt, M. R. Collins, G. M. Gallego, N. W. Sach, J. E. Spangler, H. Zhu, J. Zhu and P. S. Baran, *Science*, 2016, **351**, 241-246.
46. L. Bu, N. Zhang, S. Guo, X. Zhang, J. Li, J. Yao, T. Wu, G. Lu, J.-Y. Ma, D. Su and X. Huang, *Science*, 2016, **354**, 1410-1414.
47. D. R. Trinkle and C. Woodward, *Science*, 2005, **310**, 1665-1667.
48. A. V. Kolobov, M. Krbal, P. Fons, J. Tominaga and T. Uruga, *Nat. Chem.*, 2011, **3**, 311-316.
49. E. D. Getzoff, K. N. Gutwin and U. K. Genick, *Nat. Struct. Mol. Biol.*, 2003, **10**, 663-668.
50. S. Lüdtke, P. Neumann, K. M. Erixon, F. Leeper, R. Kluger, R. Ficner and K. Tittmann, *Nat. Chem.*, 2013, **5**, 762-767.
51. Y. S. Zholdassov, L. Yuan, S. R. Garcia, R. W. Kwok, A. Boscoboinik, D. J. Valles, M. Mariani, A. Martini, R. W. Carpick and A. B. Braunschweig, *Science*, 2023, **380**, 1053-1058.
52. M. Wuttig, D. Lüsebrink, D. Wamwangi, W. Welnic, M. Gilleßen and R. Dronskowski, *Nat. Mater.*, 2007, **6**, 122-128.
53. A. V. Kelleghan, A. S. Bulger, D. C. Witkowski and N. K. Garg, *Nature*, 2023, **618**, 748-754.
54. T. Stauch and A. Dreuw, *J. Chem. Phys.*, 2014, **140**, 134107.
55. C. E. Colwell, T. W. Price, T. Stauch and R. Jasti, *Chem. Sci.*, 2020, **11**, 3923-3930.
56. T. Stauch and A. Dreuw, *Angew. Chem., Int. Ed.*, 2016, **55**, 811-814.
57. T. Stauch and A. Dreuw, *Chem. Sci.*, 2017, **8**, 5567-5575.
58. A. Trofimova, M. Diamandas, C. Brien, N. Khasanzoda, A. J. Lough and A. K. Yudin, *J. Am. Chem. Soc.*, 2024, **146**, 23365-23375.
59. J. P. Reid and M. S. Sigman, *Nature*, 2019, **571**, 343-348.
60. M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632-672.
61. J. M. Herbert, *J. Chem. Phys.*, 2019, **151**, 170901.
62. S. Li, W. Li and J. Ma, *Acc. Chem. Res.*, 2014, **47**, 2712-2720.
63. K. Raghavachari and A. Saha, *Chem. Rev.*, 2015, **115**, 5643-5677.
64. X. He, T. Zhu, X. Wang, J. Liu and J. Z. H. Zhang, *Acc. Chem. Res.*, 2014, **47**, 2748-2757.
65. M. A. Collins and R. P. A. Bettens, *Chem. Rev.*, 2015, **115**, 5607-5642.
66. S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *J. Mol. Struct.: THEOCHEM*, 1999, **461-462**, 1-21.
67. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding and K. Morokuma, *Chem. Rev.*, 2015, **115**, 5678-5796.
68. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
69. F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.
70. C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652-1671.
71. X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, *J. Chem. Inf. Model.*, 2020, **60**, 3408-3415.
72. P. O. Dral, F. Ge, Y.-F. Hou, P. Zheng, Y. Chen, M. Barbatti, O. Isayev, C. Wang, B.-X. Xue, M. Pinheiro Jr, Y. Su, Y. Dai, Y. Chen, L. Zhang, S. Zhang, A. Ullah, Q. Zhang and Y. Ou, *J. Chem. Theory Comput.*, 2024, **20**, 1193-1213.
73. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.
74. P. Zheng, R. Zubatyuk, W. Wu, O. Isayev and P. O. Dral, *Nat. Commun.*, 2021, **12**, 7022.
75. J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192-3203.
76. C. Devereux, J. S. Smith, K. K. Davis, K. Barros, R. Zubatyuk, O. Isayev and A. E. Roitberg, *J. Chem. Theory Comput.*, 2020, **16**, 4192-4202.
77. R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.*, 2007, **79**, 291-352.
78. L. a. W. D. Schrödinger *PyMOL*, 2.4.0; 2020.
79. E. H. Krenske, E. C. Davison, I. T. Forbes, J. A. Warner, A. L. Smith, A. B. Holmes and K. N. Houk, *J. Am. Chem. Soc.*, 2012, **134**, 2434-2441.
80. R. S. Paton, S. Kim, A. G. Ross, S. J. Danishefsky and K. N. Houk, *Angew. Chem., Int. Ed.*, 2011, **50**, 10366-10368.
81. F. Schoenebeck, D. H. Ess, G. O. Jones and K. Houk, *J. Am. Chem. Soc.*, 2009, **131**, 8121-8133.

82. Z. Ma, Z. Yan, X. Li and L. W. Chung, *J. Phys. Chem. Lett.*, 2023, **14**, 1124-1132.
83. X. Li, T. Liao and L. W. Chung, *J. Am. Chem. Soc.*, 2017, **139**, 16438-16441.
84. R. Liang, Q. Zhou, X. Li, M. W. Wong and L. W. Chung, *J. Org. Chem.*, 2023, **88**, 10460-10469.
85. A. G. Green, P. Liu, C. A. Merlic and K. N. Houk, *J. Am. Chem. Soc.*, 2014, **136**, 4575-4583.
86. M. Garcia-Borras, S. B. J. Kan, R. D. Lewis, A. Tang, G. Jimenez-Oses, F. H. Arnold and K. N. Houk, *J. Am. Chem. Soc.*, 2021, **143**, 7114-7123.
87. N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046-15047.
88. K. J. Shea and J. S. Kim, *J. Am. Chem. Soc.*, 1992, **114**, 4846-4855.
89. M. Morimoto, S. M. Bierschenk, K. T. Xia, R. G. Bergman, K. N. Raymond and F. D. Toste, *Nat. Catal.*, 2020, **3**, 969-984.
90. F. H. Arnold, *Angew. Chem., Int. Ed.*, 2018, **57**, 4143-4148.
91. M. Radon, *J. Chem. Theory Comput.*, 2014, **10**, 2306-2321.
92. Z. Yan, D. Wei, X. Li and L. W. Chung, *Nat. Commun.*, 2024, **15**, 4181.