Understanding chemical reactivity using the activation strain model

Vermeeren, Pascal; van der Lubbe, Stephanie C.C.; Fonseca Guerra, Célia; Bickelhaupt, F. Matthias; Hamlin, Trevor A.

Link to publication in VU Research Portal

Understanding chemical reactivity using the activation strain model

Pascal Vermeeren [1], Stephanie C. C. van der Lubbe[1], Célia Fonseca Guerra[1,2], F. Matthias Bickelhaupt [1,3]* and Trevor A. Hamlin [1]*

Understanding chemical reactivity through the use of state-of-the-art computational techniques enables chemists to both predict reactivity and rationally design novel reactions. This protocol aims to provide chemists with the tools to implement a powerful and robust method for analyzing and understanding any chemical reaction using PyFrag 2019. The approach is based on the so-called activation strain model (ASM) of reactivity, which relates the relative energy of a molecular system to the sum of the energies required to distort the reactants into the geometries required to react plus the strength of their mutual interactions. Other available methods analyze only a stationary point on the potential energy surface, but our methodology analyzes the change in energy along a reaction coordinate. The use of this methodology has been proven to be critical to the understanding of reactions, spanning the realms of the inorganic and organic, as well as the supramolecular and biochemical, fields. This protocol provides step-by-step instructions—starting from the optimization of the stationary points and extending through calculation of the potential energy surface and analysis of the trend-decisive energy terms—that can serve as a guide for carrying out the analysis of any given reaction of interest within hours to days, depending on the size of the molecular system.

Introduction

By using computational techniques, researchers spanning all realms of chemistry now have the ability to accurately compute the energies and molecular structures of a wide variety of reactions. This equips them with the necessary tools to understand chemical reactivity, which enables them to both predict reactivity and rationally design novel reactions. In recent decades, immense progress has been made in the field of quantum chemistry[1,2], particularly in the development of density functional theory methods (DFT)[3–5]. The combination of highly accurate DFT methods and the ever-increasing computational resources readily available to scientists has led to a surge in the use of chemical modeling. Computational chemistry software enables the modeling of a large variety of chemical systems and processes with sufficient accuracy[6–9], providing an in-depth understanding of chemical reactivity[10–15]. Once the factors governing chemical reactivity are rationalized, a systematic optimization of reaction conditions and other parameters can then be carried out.

The activation strain model (ASM)[16] of reactivity is a powerful and robust method of analyzing and understanding a wide range of chemical reactions. In this fragment-based approach, the ASM relates the relative energy of a molecular complex along the reaction energy profile to two factors, namely, (i) the energy needed to deform the reactants from their optimum geometry into the geometry required to react and (ii) the actual interaction energy between these deformed reactants. This method provides the user with insight into the factors that govern the height of the reaction barrier and thus the trends in reactivity. The ASM has been the subject of multiple reviews[17–19]; however, what is ultimately missing from the literature is a step-by-step protocol that can serve as a guide for carrying out the analysis for any given reaction of interest. Although the ASM can be used in combination with many quantum chemical (QC) software packages, here, we mainly focus on the application of the ASM within the Amsterdam Density Functional (ADF)[20,21] software package. The reason is that in the case of the ADF program, the ASM can extended with (i) a matching energy decomposition analysis (EDA), which decomposes the interaction energy (obtained with the ASM) into a number of physically meaningful terms, and (ii) a molecular orbital (MO) analysis, in which

---

[1]Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [2]Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, Leiden, The Netherlands. [3]Institute for Molecules and Materials, Radboud University Nijmegen, Nijmegen, The Netherlands. *e-mail: f.m.bickelhaupt@vu.nl; t.a.hamlin@vu.nl
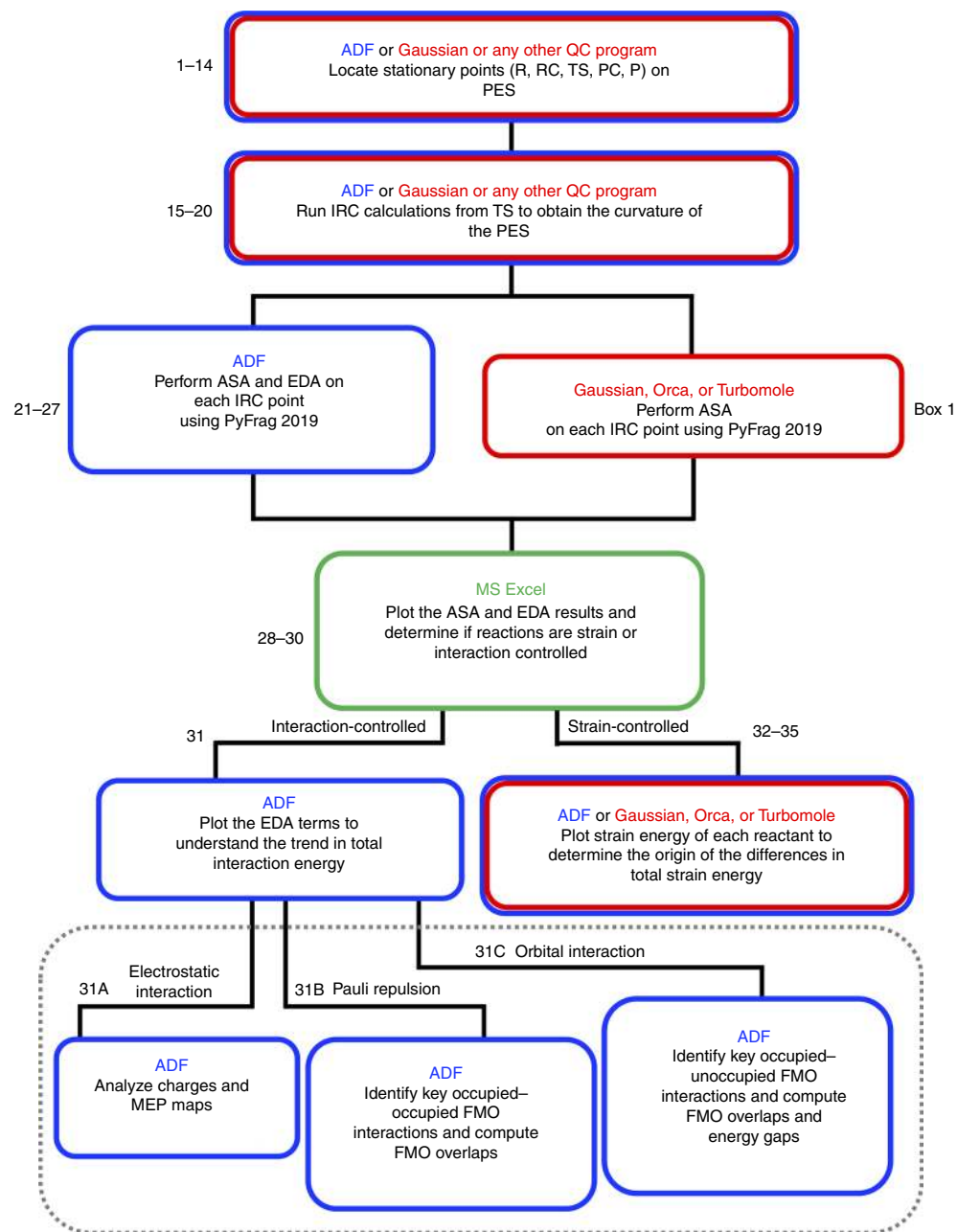
Fig. 1 | General workflow of the procedure. Schematic overview of the general workflow described in this protocol for ADF, as well as any other QC software package (Box 1). Step numbers of the Procedure are shown.

these decomposed energy terms are further rationalized by the MOs of the reactants. Both the EDA and MO analysis are powerful tools for rationalizing the factors controlling molecular reactivity[22].

Overview of the procedure

The protocol involves the following six key actions schematically shown in the workflow (Fig. 1), which is split up into 35 steps.

Locating the stationary points

First, the stationary points need to be located along the potential energy surface: this includes the reactant(s) (R), reactant complex (RC), product complex (PC), and/or product(s) (P) (Steps 1–4), as well as the transition state (TS) (Steps 5–14). The steps described in this procedure are specific to the

Box 1 | ASA with Gaussian, Orca, or Turbomole ● Timing 1–24 h, depending on the size of the molecules

Procedure
1 Install PyFrag 2019 by following the steps found in the documentation (https://pyfragdocument.readthedocs.io/en/latest/install.html#activation-strain-analysis-asa-module-of-pyfrag-2019).
2 Create a directory containing the PyFrag 2019 input file (see Supplementary Methods 6 for Gaussian, Supplementary Methods 7 for Orca, or Supplementary Methods 8 for Turbomole) and the .amv file produced by the IRC calculation.
3 Open the PyFrag 2019 input file and allocate, in the `INPUT_SPECS` block, the .amv file of the IRC calculation, using the following key:

```
output file = [filename.amv]
```

4 Specify the two interacting fragments, for example, the diene and dienophile in a Diels–Alder reaction or the catalyst and reactant in an oxidative addition reaction, with the following statement:

```
frag1 = [name_frag1]
[#atomnrs_fragment1]
end frag1
frag2 = [name_frag2]
[#atomnrs_fragment2]
end frag2
```

Here, for each fragment, the user must supply the numbers for the atoms, in a column, as they appear in the IRC file.
5 To print the strain energy, the user should specify the energies of the equilibrium geometries of the fragments.

```
print strain frag1 [equil_energy_fragment1]
print strain frag2 [equil_energy_fragment2]
```

The equilibrium energy will simply be subtracted from the energy of the respective fragment. PyFrag 2019 will print the individual strain values for each fragment, together with the total strain energy.
6 To plot the ASA data, one must select a critical reaction coordinate. Common reaction coordinates are, for example, the forming C–C bond in a Diels–Alder reaction or the dissociating C–X bond in an oxidative addition reaction.

```
print bond X Y [bond_diff]
```

Here, `X` and `Y` are the atom numbers of the atoms involved in the bond that changes during the reaction, followed by the initial bond length `[bond_diff]` between atoms X and Y. By defining the `[bond_diff]`, PyFrag 2019 subtracts this value from the actual bond length.
7 Run the PyFrag 2019 input file to obtain the ASA data, which will be printed in the resulting .txt file, with the following command.

```
pyfrag -x [gaussian/orca/turbomole] [inputfile_name].in
```

ADF software package. However, the stationary points can also be located by using any suitable QC software package.

Calculating potential energy surface
Steps 15–20 of the protocol outline how to compute the intrinsic reaction coordinate (IRC) from the TS to both the R (or RC) and P (or PC). The IRC calculation is performed to obtain the curvature of the potential energy surface (PES) of the chemical reaction of interest by following the steepest descent to the two adjacent minimum-energy structures, resulting in all energies and molecular structures between the stationary points. These steps can also be performed using any suitable QC software package.

Performing the activation strain and energy decomposition analysis
How the user performs the next steps depends on which QC software package is being used. If ADF is used, an ASA and EDA of each point along the IRC or energy profile can be performed using the provided, free-of-charge open-source PyFrag 2019[23,24] program (Steps 21–27). When QC software packages other than ADF are used, the user is currently able to perform only the ASA. PyFrag 2019 is currently compatible with Gaussian, Orca, and Turbomole, and it therefore can be used to perform the ASA with these QC programs (Box 1). The ASM decomposes the relative energy of

a molecular complex along the reaction energy profile, $\Delta E(\zeta)$, into two terms, namely the total strain energy of the reactants and their mutual total interaction energy:

$$\Delta E(\zeta) = \Delta E_{\text{strain}}(\zeta) + \Delta E_{\text{int}}(\zeta) \tag{1}$$

Here the total strain energy, $\Delta E_{\text{strain}}(\zeta)$, is the energy needed to deform the reactants into the geometry they adopt in the interacting complex and is affected by their rigidity. The total interaction energy, $\Delta E_{\text{int}}(\zeta)$, is the actual interaction between the deformed reactants and can be further decomposed by using the EDA scheme.

Plotting the data

Steps 28–30 involve the plotting of the data obtained in the previous steps. Analysis of these plots will provide the user with insights into whether the reaction under study is strain or interaction controlled. For analysis of interaction-controlled reactions, the user should use the options described in Step 31. For strain-controlled reactions, the user should follow Steps 32–35.

Analyzing the interaction energy

Step 31 describes the analysis of the interaction energy by using the canonical EDA scheme, which can be done only by using the ADF software package. Owing to the flexibility of our protocol, similar energy decomposition schemes implemented in other suitable QC programs can also be integrated into the protocol. This is, however, not discussed in the current protocol and, in order to do this, one needs sufficient knowledge of programming. The EDA decomposes the total interaction energy into the electrostatic interaction, Pauli repulsion, orbital interaction, and dispersion interaction:

$$\Delta E_{\text{int}}(\zeta) = \Delta V_{\text{elstat}}(\zeta) + \Delta E_{\text{Pauli}}(\zeta) + \Delta E_{\text{oi}}(\zeta) + \Delta E_{\text{disp}}(\zeta) \tag{2}$$

The electrostatic energy, $\Delta V_{\text{elstat}}(\zeta)$, is the classic Coulomb interaction between the unperturbed charge distributions of the deformed reactants and is usually attractive. The Pauli repulsion, $\Delta E_{\text{Pauli}}(\zeta)$, constitutes the destabilizing interaction between occupied orbitals and is responsible for any steric repulsion. The orbital interaction energy, $\Delta E_{\text{oi}}(\zeta)$, accounts for the polarization (empty–occupied orbital mixing on one fragment because of the presence of another fragment) and charge transfer (donor–acceptor interactions between occupied orbitals on one fragment and unoccupied orbitals on the other, including the HOMO–LUMO (highest occupied MO–lowest unoccupied MO interactions). The dispersion energy, $\Delta E_{\text{disp}}(\zeta)$, accounts for the long-range dispersion interactions when an explicit dispersion correction is included.

Each of these individual energy terms can be further analyzed as follows. For $\Delta V_{\text{elstat}}(\zeta)$, analyze atomic charges (e.g., Voronoi deformation density (VDD)[25], Hirshfeld[26], or multipole-derived[27] charges) and molecular electrostatic potential (MEP) distributions on each reactant (Step 31A). For $\Delta E_{\text{Pauli}}(\zeta)$, perform a Kohn-Sham MO analysis by analyzing the overlap between the occupied MOs on each reactant (Step 31B). For $\Delta E_{\text{oi}}(\zeta)$, perform a Kohn–Sham MO analysis by comparing the orbital overlap and energy difference between the occupied and unoccupied fragment MOs (FMOs) of the reaction of interest (Step 31C).

Analyzing the strain energy

If the activation strain diagram reveals that the reaction is strain controlled, the total strain energy should be decomposed into the strain associated with deforming each respective reactant (Steps 32–35):

$$\Delta E_{\text{strain}}(\zeta) = \Delta E_{\text{strain,frag1}}(\zeta) + \Delta E_{\text{strain,frag2}}(\zeta) \tag{3}$$

Here, $\Delta E_{\text{strain,frag1}}(\zeta)$ is the strain energy corresponding to the deformation of fragment 1, and $\Delta E_{\text{strain,frag2}}(\zeta)$ is the strain energy corresponding to the deformation of fragment 2. Comparing the fragment strain energies, which originate from the rigidity of the fragments, with their corresponding geometrical deformations will help protocol users to determine and understand the origins of the differences in total strain energy.

Applications of the method

The ASM can be applied to all unimolecular and bimolecular reactions in both homogeneous and heterogeneous systems and has been used routinely by theoretical and experimental chemists[28–42]. We provide specific examples of the ASM being applied to understand inorganic, organic, and
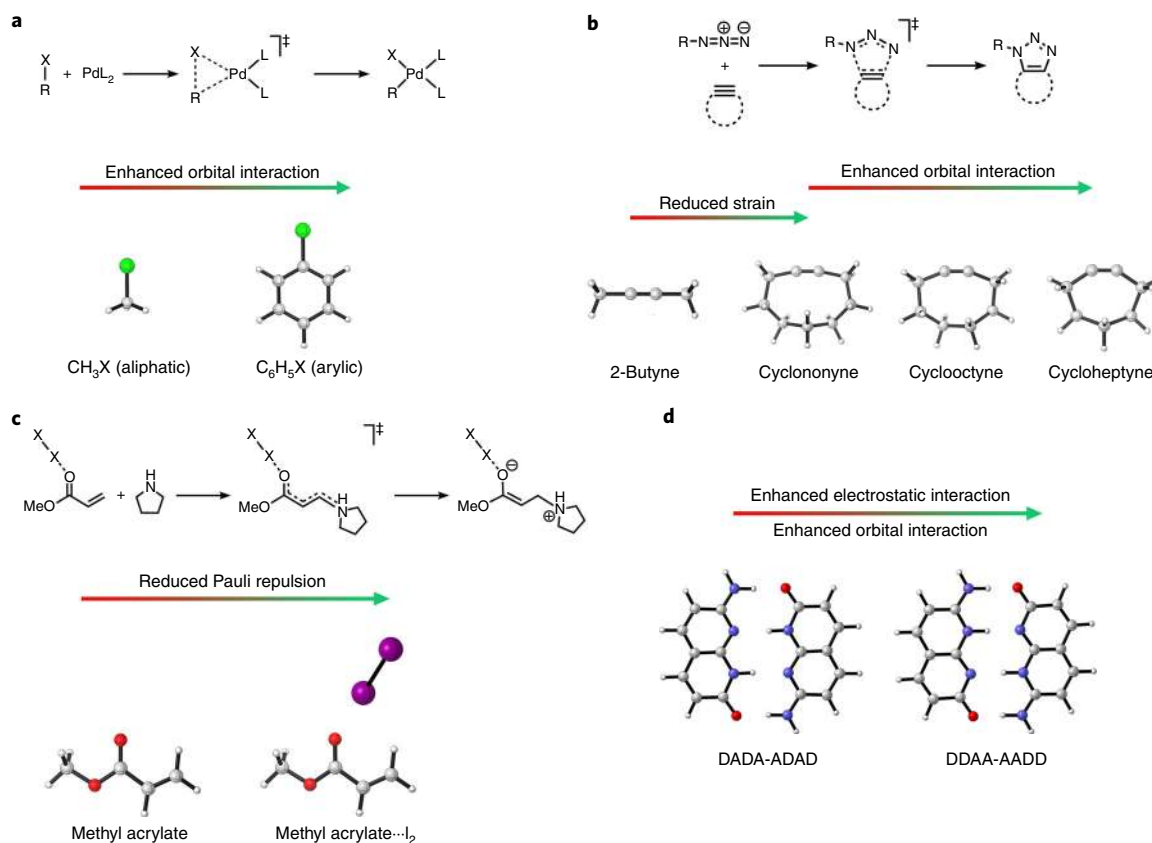
Fig. 2 | Examples of systems for which the ASM has been successfully applied. a, The oxidative addition of C–X bonds. **b**, Reactivity of cycloalkynes in 1,3-dipolar cycloadditions. **c**, Dihalogen-catalyzed Michael additions. **d**, Bonding mechanism in hydrogen-bonded systems. **a** adapted from ref. [43] under a Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/legalcode). **b** adapted from ref. [44] under a Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/legalcode). **c** adapted with permission from ref. [45], Wiley. **d** adapted with permission from ref. [47], American Chemical Society.

supramolecular chemistries, namely, the transition metal–mediated oxidative addition of C–X bonds in cross-coupling reactions, the reactivity of cycloalkynes in 1,3-dipolar cycloadditions, the reactivity of dihalogen-catalyzed Michael addition reactions, and the bonding mechanism in hydrogen-bonded systems.

Previously, the ASM has been used to explain why the reaction barriers of the oxidative insertion of a palladium catalyst into arylic C–X bonds (X = H, Cl, CH_3) is consequently lower than those for the corresponding aliphatic C–X bonds (Fig. 2a; ref. [43]). The in-depth quantitative analysis revealed that the trend in barrier heights is controlled by the interaction energy, which, with the help of the EDA, could be attributed to the orbital interaction. The differences in orbital interaction between the activation of arylic and aliphatic C–X bonds were rationalized by performing a Kohn–Sham MO analysis, which revealed that the HOMO–LUMO energy gap of the former is substantially smaller than that of the latter, which results in more stabilizing orbital interactions.

In addition to inorganic reactions, the ASM has also been used for analyzing organic cycloaddition reactions, such as the reactivity of cycloheptyne, cyclooctyne, and cyclononyne in the strain-promoted azide–alkyne cycloaddition (SPAAC) with methyl azide (Fig. 2b; ref. [44]). The ASM, in combination with the EDA, revealed that, in contrast to the widespread belief, the increased degree of geometrical pre-distortion when the ring size of the substrate is reduced, effectively leads to enhanced reactivity due to not only reduced strain energy but also more-stabilizing orbital interaction. By studying the FMOs using a Kohn–Sham MO analysis, the trend in orbital interactions could be ascribed to both a smaller HOMO–LUMO energy gap and a larger orbital overlap.

Another interesting and well-known organic reaction that was recently studied with the ASM and EDA methods is the dihalogen catalyzed aza-Michael addition of pyrrolidine and methyl acrylate (Fig. 2c; ref. [45]). The findings of this study revealed that the activation barriers decrease when the reaction is catalyzed by heavier dihalogens ($F_2 < Cl_2 < Br_2 < I_2$). The activation strain and bonding

analysis demonstrated that an unexpected physical factor controls the computed reactivity trends, namely, the Pauli repulsion between the nucleophile (pyrrolidine) and the Michael acceptor (methyl acrylate). With the aid of a Kohn–Sham MO analysis, this behavior was found to originate from coordination of the dihalogen to the carbonylic oxygen of the Michael acceptor, which, in turn, polarizes the conjugated π system away from the electrophilic carbon atom and results in a reduced Pauli repulsion toward the lone pair of the incoming nucleophile.

The ASM can also be extended to any point along the energy profile. For example, it can be used to scan the energy profile along the point of equilibrium of a complex to pinpoint the driving force behind the trends in geometry and stability for hydrogen-bonded systems (not included in the Procedure; Fig. 2d). The advantage of this approach is that hydrogen-bonded dimers can be compared with identical bond lengths, which enables one to distinguish the energy terms that are effectively stronger, and are thus responsible for the trends in geometry and stability, from the terms that are simply enhanced because of the shorter hydrogen bond length. This approach has been used successfully on mismatched DNA base pairs, revealing that steric repulsion can be a decisive factor for relative hydrogen bond strengths and lengths[46]. In addition, this procedure has been used on quadruple hydrogen–bonded dimers to rationalize why and how the secondary electrostatic interaction model has been predictive[47].

Limitations

A potential limitation of the protocol is that, in its current form, the only EDA scheme that it interfaces with is the one in the ADF program. All the other steps can be carried out with any other suitable QC program. Importantly, owing to the flexibility of the protocol, it can be adapted to use any interaction energy decomposition method included in other suitable QC software packages. A second limitation is that the EDA scheme is not yet compatible with implicit solvation (all other steps can be carried out in implicit solvation). The EDA in implicit solvation is currently being implemented by our groups. There is, however, a workaround in which one decomposes the solution-phase energy profile into two terms, namely, the solute and solvation energies[48–51]. The user can perform the standard (i.e., gas-phase) EDA on the solute, which has the geometry as obtained by the geometry optimization with implicit solvation. Recently, solvation has been shown to alter the EDA results, including screening of the dispersion interaction by about 70%; thus, care should be exercised when carrying out this analysis[52]. Finally, one might perceive unimolecular reactions as a possible limitation because of the absence of two clearly defined reactants. However, by using a careful and chemically meaningful fragmentation scheme, unimolecular reactions can easily be studied with the same analysis tools as used for bimolecular reactions[17,19].

Comparison with other methods

A strongly related method involves performing a single-point ASA on a stationary point, such as the TS, or the equilibrium structure of a molecular complex instead of along the entire reaction coordinate. This alternative method is prevalent in the literature[53,54], probably because it is 'cheap' from a computational perspective. We must stress the fact that analysis of the strain and interaction energy at a single stationary point should be interpreted carefully, because these energy components are highly dependent on the position of this point on the reaction coordinate/potential energy surface[19,55]. To illustrate this point, we consider the activation strain diagrams of two generic reactions A and B (Fig. 3). Reaction B clearly proceeds with a lower activation barrier and thus the reaction is faster than reaction A. The single-point analysis at the TS structures would suggest that reaction B proceeds with a lower barrier solely due to a lower strain energy. In fact, the interaction energy at the TS, for reactions A and B, does not differ at all. Now, when considering the analysis along the entire reaction coordinate, as follows from the ASM, it becomes clear that the interaction energy of reaction B is, at every point along the reaction coordinate, more stabilizing than that of reaction A, while the two strain energies are nearly identical. Thus, two disparate views of the factors controlling the reactivity emerge from the two approaches when the TS structures of A and B occur at different points on the reaction coordinate. Therefore, it is clear that the user should exhibit caution when comparing the strain and interaction energies of different reactions with transition states (TSs) occurring at different points along the reaction coordinate. There is, however, an acceptable application of the analysis of the strain and interaction energy at a single point, namely, when these points are located at a consistent geometry along the reaction coordinate (vertical line in Fig. 3). If one were
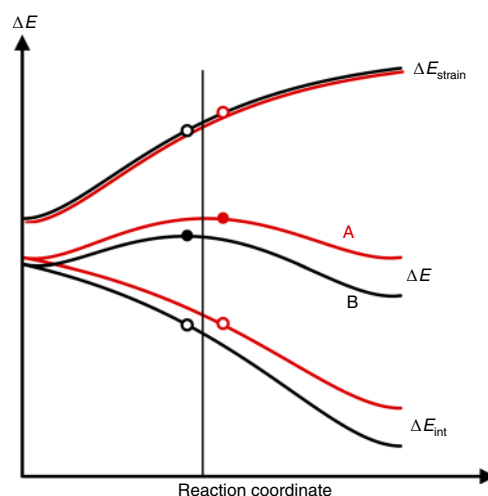
Fig. 3 | Activation strain diagram. Two generic reactions A (red curves) and B (black curves), in which the positions of the TSs are indicated with dots and the consistent geometry along the reaction coordinate with a vertical line. Image adapted with permission from ref. [19], Royal Society of Chemistry.

to perform this type of analysis, the position of the consistent geometry on the reaction coordinate should be explicitly mentioned.

Experimental design

The accuracy of a density functional theory (DFT) calculation is determined by, among others, the exchange-correlation (XC) energy functional and the so-called basis set, which consists of atom-centered functions that are used to construct the orbitals and, thus, the electron density. Therefore, before carrying out the following procedure, one must carefully select both the XC functional and the basis set. There are many different XC functionals available, and which one is the most suitable depends on the reaction that is being studied. Generally, larger basis sets (i.e., sets with a higher number of atomic basis functions) give more accurate results. In this section, we suggest possible combinations of XC functionals and basis sets that have proven robust and provide accurate results in combination with the protocol. We use the standard notation 'XC functional/basis set'. So, for example, 'BLYP/TZ2P' refers to the BLYP functional in combination with a TZ2P basis set.

For investigations involving the analysis of catalytic cycles, for example, the oxidative insertion of a closed-shell palladium catalyst into a C–X bond, the relativistic corrected ZORA-BLYP/TZ2P level will give accurate results. This approach was extensively tested against ab initio reference benchmarks from hierarchical series up until CCSD(T)[56–60]. The D3 dispersion correction[61], in combination with Becke–Johnson (BJ) damping[62,63], should be used when dispersion interactions are suspected to be important. When studying nucleophilic substitution reactions, use of the OLYP/TZ2P level, in combination with the ZORA relativistic correction, depending on the size of the atoms involved in the reaction, is suggested[64,65]. For the analysis of other organic reactions, such as the Diels–Alder reaction or 1,3-dipolar cycloadditions, we recommend the use of either the M06-2X or BP86 XC functional in combination with the TZ2P basis set[55,66]. When dispersion interactions are expected to play an important role in these organic reactions, Hamlin et al.[55] suggest the use of BLYP-D3(BJ)/TZ2P in order to obtain accurate and reliable results. Finally, for studying the molecular complexation based on non-covalent interactions, such as hydrogen-bonded arrays, we recommend using BLYP-D3(BJ)/TZ2P in analogy with the studies performed by our group[46,47,50,51].

Materials

- Computer (Linux, Windows, or Macintosh)
- Computer cluster (Intel i5 or better 64-bit CPU, 8 GB RAM, 250 GB SSD)
- Python 3 package (https://www.python.org/download/releases/3.0/)
- QC software package for locating the stationary points and computing the IRC, such as Amsterdam Density Functional 2018 (ADF; https://www.scm.com), Gaussian (https://gaussian.com), Orca (https://orcaforum.kofo.mpg.de/app.php/portal), or Turbomole (http://www.turbomole.com)

- The PyFrag 2019 package, in combination with ADF2018, Gaussian, Orca, or Turbomole, for performing the activation strain method (ASM) calculations (https://pyfragdocument.readthedocs.io/en/latest/install.html#activation-strain-analysis-asa-module-of-pyfrag-2019, https://github.com/sunxb05/PyFrag)
- The PyFrag 2019 package in combination with ADF2018 for performing the EDA calculations (https://pyfragdocument.readthedocs.io/en/latest/install.html#activation-strain-analysis-asa-module-of-pyfrag-2019, https://github.com/sunxb05/PyFrag)
- Excel or any other spreadsheet program to plot the energy terms as a function of the reaction coordinate (https://www.microsoft.com/en-us/p/excel/cfq7ttc0k7dx?activetab=pivot%3aoverviewtab)

Procedure

Locating reactant(s), reactant complex, product complex, and/or product(s) on the PES
● **Timing** minutes to days, depending on the size of the molecules

1 Install ADF, following the steps found on the installation page of the SCM website (https://www.scm.com/doc/Installation/Installation.html?highlight=download). Double-click on the ADF icon on your desktop to launch ADFjobs. Go to *SCM → New Input* to open ADFinput.

2 Draw the molecule/import coordinates from the literature. Go to *Details → Run Script* to copy the coordinates and paste them into the Geo_Freq input file (see Supplementary Methods 1).

3 Run the geometry optimization together with a frequency calculation.

4 Open the generated .t21 file using ADFspectra to see if the optimized structure is an energy minimum, that is, frequencies should be positive and all peaks of the IR spectrum should point in the same direction.
   ▲ **CRITICAL STEP** If one or more imaginary frequencies are calculated in the IR spectrum, the stationary point is not an energy minimum.
   ▲ **CRITICAL STEP** The calculated energy minimum can be either a local or global minimum. Users should attempt to locate the global minimum using their chemical intuition by computing various possible conformations (the global minimum structure will have the lowest absolute energy). When working with flexible molecules with many degrees of freedom, users can consider using an automated molecular mechanics or GFN2-xTB[67] approach to evaluate the relative energy of relevant conformers, but these techniques will not be further discussed here.
   ? TROUBLESHOOTING

Locating the TS on the PES ● **Timing** minutes to days, depending on the size of the molecules

5 Launch ADFinput.

6 Copy/paste the coordinates of the product (or product complex) into ADFinput.

7 Locate the bonds that change over the course of the reaction, for example, the newly formed C–C bonds in a Diels–Alder reaction or the dissociation of a C–X bond by oxidative addition.

8 Set up the geometry constraints by changing the following line in the linear_transit input file (see Supplementary Methods 2):

```
CONSTRAINTS
DIST X Y start=A end=B
END
```

   X and Y are the atom numbers of the atoms involved in the bond that changes during the reaction (located in step 7). A and B are the initial and final X–Y bond distances, i.e., the distance in the respective product (or product complex) and reactant (or reactant complex). It is possible to constrain multiple bond distances by simply adding an additional `DIST X Y start=A end=B` line.

9 Set up the number of linear transit steps by changing the following line in the geometry block of the linear_transit input file:

```
Geometry
lineartransit [#steps]
End
```

Here, `#steps` is the number of linear transit steps; 9 linear transit steps are usually sufficient to obtain an approximate potential energy surface. ADF will optimize the molecule at each step along the reaction coordinate while retaining the constraint(s).

10   Run the LT in order to simulate the reaction pathway within the chosen geometry constraints.

11   Open the resulting .t21 file using ADFmovie to locate the approximate geometry of the TS. Go to *Graph → Energy* to display the simulated PES. Go to *View → Converged geometries only* in order to display only the converged geometries. The highest point of the energy graph is the geometry that most closely resembles the TS. Save this geometry by clicking on this point in the energy graph, followed by selecting *File → Save geometry*. This will generate a .txt file containing the coordinates.

12   Copy the coordinates from the .txt file of Step 11 to the TSRC input file (see Supplementary Methods 3) and specify a reaction coordinate along which the TS will be searched for:

```
TSRC
DIST X Y [fac]
END
```

Here, `X` and `Y` are the atom numbers of the atoms involved in the bond that changes during the reaction, which most often is the same bond specified as in Step 7. These are followed by a factor `[fac]`, which can be either 1.0 or –1.0, defining an increase or decrease of the X–Y bond distance, respectively. It is possible to specify multiple reaction coordinates by simply adding an additional `DIST X Y [fac]` line.

13   Run the TSRC together with a frequency calculation to find the TS.

14   Open the generated .t21 file, using ADFspectra, to check whether an actual TS was located. There should be a single peak with an imaginary frequency, that is, one peak that is pointing downward. Click on this peak to see the corresponding normal mode, which should be the mode of the reaction.
▲ CRITICAL STEP   If more than a single imaginary frequency in the IR spectrum is computed, then the stationary point is a higher-order saddle point and you therefore cannot continue with Step 15. Refer to the Troubleshooting section and repeat the calculations from Step 12.
? TROUBLESHOOTING

Running IRC calculations from the TS to obtain the curvature of the PES ● Timing minutes to days, depending on the size of the molecules

15   Import the coordinates of the TS into the IRC input file (see Supplementary Methods 4).

16   Set up the IRC calculation by adjusting the following line in the input file (see Supplementary Methods 4):

```
Geometry
IRC Backward=[True/False] Forward=[True/False] POINTS=A STEP=B
End
```

Here, `A` is the maximum number of IRC points computed in the run (ranging from 0 to 99), and `B` is the step length when proceeding from one IRC point to another along the path (ranging from 0 to 1). The default step length is 0.2 $(amu)^{1/2}$ Bohr (amu is atomic mass unit), which will give you an accurate description of the PES. By specifying `Backward` or `Forward` in the geometry block, you can run the IRC calculations parallel to the adjacent minima, meaning that the pathways from the TS to the reactant (or reactant complex) and to the product (or product complex) are calculated separately. When no direction is specified, the backward and forward reaction pathways will subsequently be computed in one job.
▲ CRITICAL STEP   Choosing to run the forward and backward IRC calculations separately obligates you to perform two PyFrag calculations in Step 23: one from the TS to the reactant (or reactant complex) and one from the TS to the product (or product complex).

17   It can be useful to restart the IRC from the Hessian that was calculated during the TSRC calculations because this Hessian will guide the point-by-point optimizations along the IRC path and the eigenvector of the lowest Hessian eigenvalue will define the initial direction of the IRC path. Locate the .t21 file of the TSRC calculation by navigating to the respective .t21 file using the following line in the IRC input file:

```
Restart [pathname]/TSRC.t21
```

18   Run the IRC calculation to obtain the PES.

19   Check the generated .out file to see if the IRC calculation has reached the minimum. The complete PES is calculated when the .out file ends with the statement 'finishing IRC path Backward/Forward'. If this statement is not present in the .out file, you should continue to perform IRC calculations, restarting from the .t21 file of the previous IRC calculation until this statement is present.

20   Confirm that your calculated reaction pathway has finished at the correct reactant (or reactant complex) and product (or product complex) by comparing the final structures of your backward and forward IRC calculation to the optimized structures of Steps 1–4.

▲ CRITICAL STEP   The user should verify that the last point of the IRC is close in both geometry and energy (ideally <1.0 kcal mol^{-1}) to the computed stationary points on either side of the TS structure. Further analysis should be performed on the complete PES, because analysis of an incomplete PES might result in skewed conclusions.

? TROUBLESHOOTING

Performing the ASA and EDA on each IRC point ● Timing 1–24 h, depending on the size of the molecules

21   Install PyFrag 2019 as described in the documentation (https://pyfragdocument.readthedocs.io/en/latest/install.html#activation-strain-analysis-asa-module-of-pyfrag-2019).

22   Create a directory containing the PyFrag 2019 input file (see Supplementary Methods 5) and the .t21 file of the backward or forward IRC calculation.

23   Open the PyFrag 2019 input file and allocate, in the PyFrag block, the .t21 file of the IRC calculation, using the following key:

```
irct21 [pathname]/IRC.t21
```

▲ CRITICAL STEP   Choosing to run the forward and backward IRC calculations separately in Step 16 obligates you to perform two PyFrag calculations: one from the TS to the reactant (or reactant complex) and one from the TS to the product (or product complex).

24   Specify the two interacting fragments, for example, the diene and dienophile in a Diels–Alder reaction or the catalyst and reactant in an oxidative addition reaction, with the following statement:

```
fragment [#atomnrs_fragment1]
fragment [#atomnrs_fragment2]
```

Here, the user must supply, for each fragment, a list of the numbers for the atoms as they appear in the IRC file.

25   To print the strain energy, the user should specify the energies of the equilibrium geometries of the fragments, which are calculated in Steps 1–4.

```
strain [equil_energy_fragment1]
strain [equil_energy_fragment2]
```

The equilibrium energy will simply be subtracted from the energy of the respective fragment. PyFrag 2019 will print the individual strain values for each fragment, together with the total strain energy.

26   To plot the ASA and EDA data, one must select a critical reaction coordinate. Common reaction coordinates are, for example, the forming C–C bond in a Diels–Alder reaction or the dissociating C–X bond by an oxidative addition reaction.

```
bondlength X Y [bond_diff]
```

Here, X and Y are the atom numbers of the atoms involved in the bond that change during the reaction (most often these are the same atoms specified in Step 12), followed by the initial bond length [bond_diff] between atoms X and Y. By defining the [bond_diff], PyFrag 2019 subtracts this value from the actual bond length.

27 Run the PyFrag 2019 input file to obtain the ASA and EDA data, which will be printed in the resulting .txt file, with the following command:

```
pyfrag [inputfile_name].in
```

Plotting ASA and EDA results ● Timing 0.5–1 h

28 Copy the data from the .txt file generated in Step 27 and paste it into a Microsoft Excel spreadsheet.

29 Plot two 'Scatter with Smooth Lines' diagrams. The first diagram will be the activation strain diagram containing the total energy (ΔE), total interaction energy (ΔE_{int}), and the total strain energy (ΔE_{strain}) as y-axis data. The second diagram is the EDA diagram, which includes the total interaction energy (ΔE_{int}), electrostatic interaction (ΔV_{elstat}), Pauli repulsion (ΔE_{Pauli}), orbital interaction (ΔE_{oi}), and, if an explicit dispersion correction is included, the dispersion interaction (ΔE_{disp}) as y-axis data. The energy values of both diagrams should be projected onto the critical reaction coordinate (x axis) that was selected in Step 26.
▲ CRITICAL STEP One should project the data onto various reaction coordinates to confirm that the trend of the energy curves is not simply an artifact of the chosen reaction coordinate.

30 Inspect the first diagram as obtained in Step 29 and identify the term(s) that govern(s) the trend in total energy ΔE. This can be done by comparing the strain energy curves for the reactions of interest and by comparing the interaction energy curves for the reactions of interest. For example, if the trend in total energy ΔE is determined entirely by the interaction energy, the strain energy curves will be superimposed or show a trend opposite to that of the total energy curves and non-overlapping interaction energy curves that follow the same trend as the total energy curves. If the reaction is interaction controlled, continue with Step 31 and skip Steps 32–35. If the reaction is strain controlled, continue with Steps 32–35. If the reaction is controlled by both the interaction energy and strain energy, continue with Steps 31–35.
▲ CRITICAL STEP Failure to identify the term that rules the trend in total energy will result in incorrect conclusions.

Analyzing ASA and EDA results ● Timing 1–24 h, depending on the size of the molecules

31 Analyze the EDA data (second graph obtained in Step 29) and identify which EDA term governs the trend in interaction energy. This term should be further analyzed to establish a physical mechanism to explain the reactivity trends of the studied reactions. If the reactivity of the reaction is controlled by the electrostatic interactions, follow option A. If the reactivity of the reaction is controlled by the Pauli repulsion, follow option B. If the reactivity of the reaction is controlled by the orbital interactions, follow option C.

(A) **Analysis of electrostatic interaction–controlled trend**

(i) Use ADFview to open both of the .t21 files of the isolated fragments (not the complex), for example, frag1.000.t21 and frag2.000.t21, generated by PyFrag 2019.

(ii) For the atomic charges, go to *Properties → Atom info*. A variety of charges can be selected here, including VDD, Hirshfeld, and multipole-derived charges. The charges can be visualized per atom by clicking on the method of choice, for example, *VDD Charge → Show*.

(iii) For the MEPs, go to *Add → Isosurface: Colored*. In the left field at the bottom, select *Density → SCF*. In the right field at the bottom, select *Potential → Coulomb Potential SCF*. The quality of the image can be improved by selecting a finer grid in *Fields → Grid* (the default grid is coarse). Check the bar option at the bottom right to see the color bar; click on the resulting color bar to open more detailed settings, including the colormap *Rainbow* option.
▲ CRITICAL STEP When comparing two or more MEPs, it is essential that their isosurface and field values be identical. Ideally, the color bar will be symmetric (i.e., the absolute min and plus field values are identical) so the green color corresponds to neutral regions.

(iv) Rationalize the differences in electrostatic interaction by comparing the charges of the atoms and/or by comparing the charge accumulation around the atoms that are directly involved in the reaction. A larger (accumulation of) charge will, in principle, result in stronger electrostatic interactions.

(B) **Analysis of Pauli repulsion–controlled trend**

(i) Use ADFview to open the .t21 files of the separated fragments generated by PyFrag 2019.

(ii) Select *Add → Isosurface: Double (+/-)* and click, in the left field at the bottom, on *Select Field → Orbitals (occupied)* in order to visualize the spatial extent of the occupied orbitals.

    (iii) Analyze which occupied fragment orbitals might overlap and might thus be involved in Pauli repulsive interactions.

    (iv) Redo the PyFrag 2019 calculation, including the following line in the PyFrag block in order to print the overlap between the orbitals that were pinpointed in Step 31B(iii):

```
overlap frag1 [MO_frag1] frag2 [MO_frag2]
```

        Here, `[MO_frag1]` and `[MO_frag2]` are the occupied orbitals of fragment 1 and fragment 2, respectively. For example, HOMO, HOMO−1, or HOMO−11.

    (v) Rationalize the differences in Pauli repulsion by comparing the predominant overlaps between the occupied orbitals of all reactions of interest, together with the spatial extent of these fragment orbitals. A larger occupied–occupied orbital overlap will result in a more destabilizing Pauli repulsion.

(C) **Analysis of orbital interaction–controlled trend**

    (i) Use ADFview to open the .t21 files of the separated fragments, generated by PyFrag 2019.

    (ii) First, go to *Add → Isosurface: Double (+/-)* and click, in the left field at the bottom, on *Select Field → Orbitals (occupied/virtual)* in order to visualize the spatial extent of the HOMOs (occupied)/LUMOs (virtual).

    (iii) Analyze which occupied/virtual fragment orbitals might overlap and be involved in the HOMO–LUMO interaction.

    (iv) Open the .t21 file of the fragment analysis calculation using ADFlevels, and consider which two-electron–two-orbital interactions between the two fragments are important. By placing the mouse on a fragment orbital, its gross population will appear. For occupied orbitals, a gross population <2.00 indicates a loss of electrons, whereas, for virtual orbitals, a gross population >0.00 indicates a gain of electrons upon complexation of the fragments. Important fragment orbitals are those for which the gross population value increases (virtual orbitals) or decreases (occupied orbitals) considerably.

    (v) Redo the PyFrag 2019 calculation, including the following lines in the PyFrag block in order to print the mutual overlap between the orbitals, as well as the orbital energy of the fragment orbitals that are included in the HOMO–LUMO interaction allocated in Step 31C(iv):

```
overlap frag1 [MO_frag1] frag2 [MO_frag2]
orbitalenergy frag1 [MO_frag1]
orbitalenergy frag2 [MO_frag2]
```

        Here, `[MO_frag1]` and `[MO_frag2]` are the occupied and/or unoccupied orbitals of fragment 1 and fragment 2, respectively. For example, HOMO, HOMO−1, or HOMO−11 and LUMO, LUMO+2, or LUMO+9.

    (vi) Rationalize the differences in orbital interaction by comparing the predominant HOMO–LUMO overlaps together with the HOMO–LUMO orbital energy gaps of all reactions of interest. A larger orbital overlap, as well as a smaller fragment orbital energy gap, will result in a more stabilizing orbital energy.

        ? TROUBLESHOOTING

Analyze strain energy results ● Timing 1–24 h, depending on the size of the molecules

32   Plot a "Scatter with Smooth Lines" diagram, that includes the following energy terms: total strain energy (ΔE_{strain}), strain energy of fragment 1 ($\Delta E_{strain,frag1}$), and strain energy of fragment 2 ($\Delta E_{strain,frag2}$), projected onto a reaction coordinate that the user selected in Step 26.

33   Compare the fragment strain energy curves and pinpoint which fragment is responsible for the trend in total strain energy.

34   Use ADFview to open the .t21 files of the reactant, as well as of some points along the reaction coordinate of the fragment allocated in Step 33.

35   Rationalize the differences in fragment strain energy by comparing the geometrical deformations of the fragments, along the reaction coordinate, with their ground state (reactant) geometry.

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1 | Troubleshooting table

Step	Problem	Possible reason	Solution
4	The reactant (complex) or product (complex) has one or more imaginary frequencies	The geometry optimization did not converge to an energy minimum	Repeat the geometry optimization (Step 3), using a modified initial structure
14	The located TS does not have any imaginary frequencies	The initial structure was not close enough in geometry to the actual TS, and thus the TS calculation has converged into an energy minimum	Repeat the TS calculation (Step 12) and change the sign of the TSRC factor
			Repeat the TS calculation (Step 12) with a different reaction coordinate
			Repeat the TS calculation (Step 12), using a modified initial structure
20	Final structures of IRC calculation do not correspond to the reactant (complex) and product (complex)	The TS used in the IRC calculations does not connect the reactant (complex) with the product (complex)	Locate the correct TS by repeating the linear transit in Steps 8–10 with different geometry constraints
	The IRC calculation is not completed within the provided wall-time	Too many steps and/or a too-small step length are specified in the input file	Reduce the number of IRC steps and/or increase the step length (Step 16)
	The IRC calculation does not work	The potential energy surface is too shallow	Reduce the size of the step length; perform a constrained optimization by means of a linear transit (Steps 8–10)
31C(vi)	The trend in orbital interaction cannot be ascribed to one term	Both the HOMO–LUMO overlap and the corresponding orbital energy gap are contributing to the trend in orbital interaction	Calculate the orbital stabilization due to the HOMO–LUMO interaction by $S^2/\Delta\varepsilon^1$, which is the orbital overlap squared divided by their difference in orbital energy, in order to explain the trend in orbital interaction

Timing

Steps 1–14, finding the stationary points on the PES: minutes to days, depending on the size of the molecules

Step 15–20, running the IRC calculations from the TS to obtain the curvature of the PES: minutes to days, depending on the size of the molecules

Step 21–27, performing ASA and EDA on each IRC point: 1–24 h, depending on the size of the molecules

Step 28–35, plotting and analyzing ASA and EDA results: 1.5–25 h, depending on the size of the molecules

Box 1, ASA with Gaussian, Orca, or Turbomole: 1–2 h, depending on the size of the molecules

Anticipated results

The results of the ASA and EDA (ASM–EDA) greatly depend on the reaction that is being studied. In the following, analysis of two representative and fundamental reactions are provided, namely, (i) the oxidative insertion of a palladium catalyst into the C–Cl bond of chloromethane and (ii) strain-promoted azide–alkyne cycloaddition. The former serves as a fundamental example illustrating how to run the complete protocol and is accompanied by all necessary input files to reproduce the results (see Supplementary Methods 9–17). The latter example is more advanced and is provided to illustrate how to analyze and compare two related chemical reactions.

The oxidative insertion of a palladium catalyst into a C–X bond, such as the C–Cl bond of chloromethane, is the rate-determining step in the Suzuki cross-coupling reaction and is, therefore, of great interest for any organic and inorganic chemist[43]. The ASM–EDA procedure, as described in this protocol, can be applied to analyze and understand this reaction step. First, one should optimize the geometries of the reactants (R), as well as those of the reactant complex (RC) and product complex
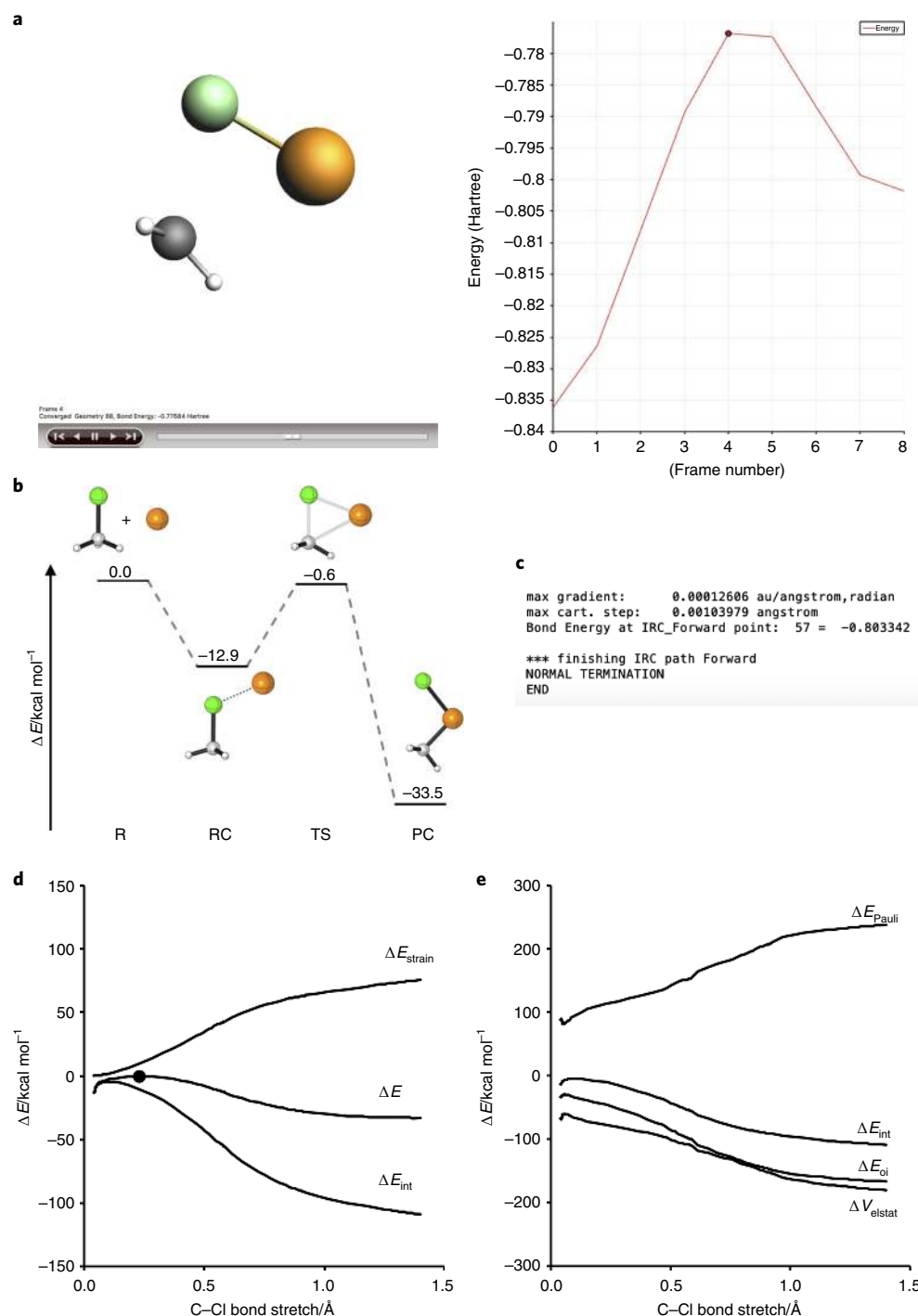
Fig. 4 | Analysis of the oxidative insertion of palladium into the C–Cl bond of chloromethane. a, Screenshot of ADFmovie, displaying the result of a linear transit calculation in which the selected, highest, point is an approximation of the TS. **b**, Relative energy profile showing the stationary points. **c**, Statement in output file of IRC forward calculation indicating that the IRC calculation reached its adjacent minimum. **d**, ASA. **e**, EDA diagram of the oxidative insertion of Pd into the C–X bond of chloromethane. Image adapted from ref. [43] under a Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/legalcode).

(PC) (Steps 1–4). This can be carried out using the input files provided in Supplementary Methods 9–12. The TS can be located by first running a linear transit calculation from the PC to the RC (Steps 5–11), where the constraints are the shortening of the C•••Cl distance (from 3.22 Å to 1.86 Å) and a sequential elongation of the C•••Pd distance (from 2.01Å to 3.86 Å) (Supplementary
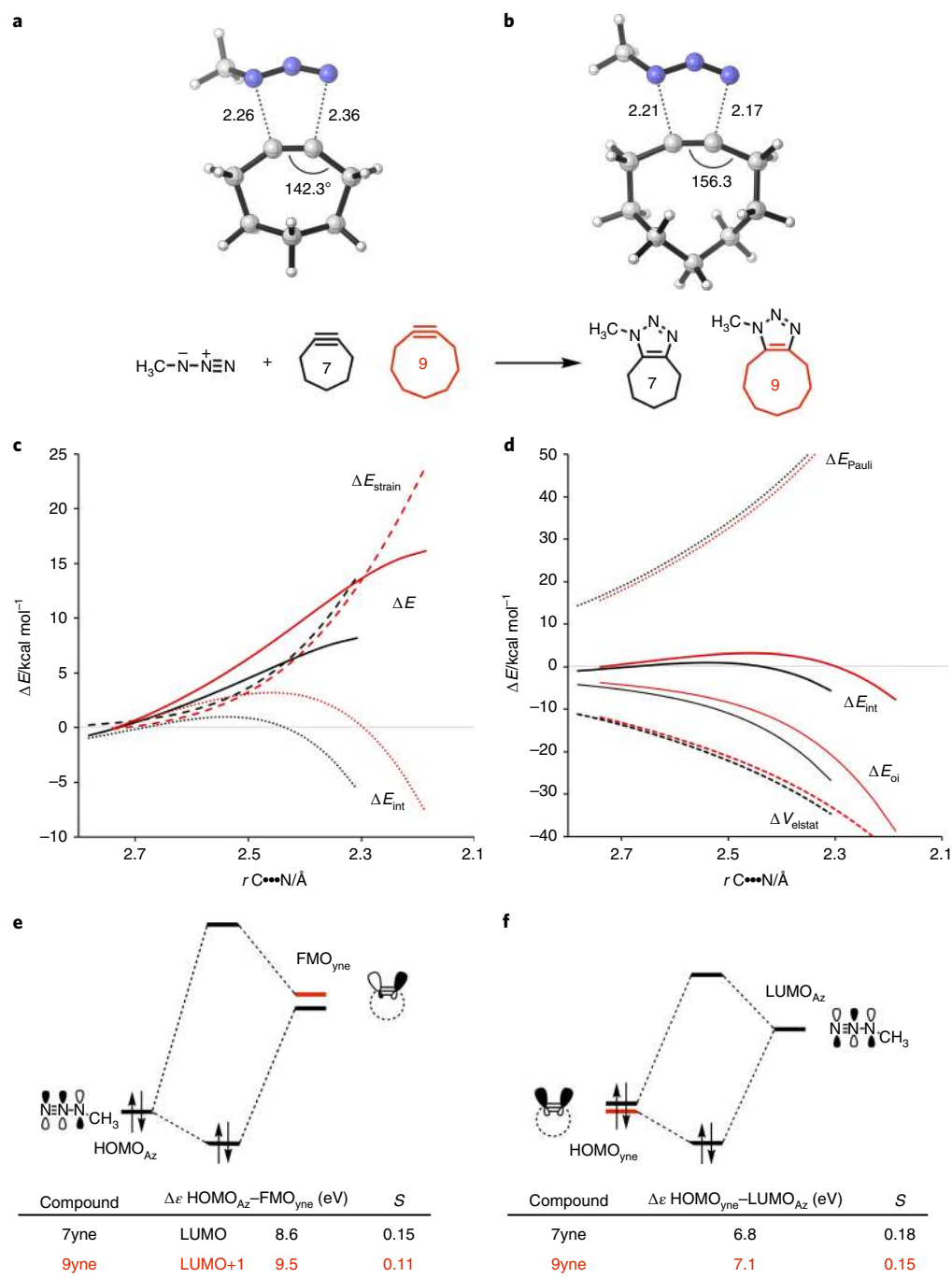
Fig. 5 | Comparison of the strain-promoted azide–alkyne cycloaddition reactivity between methyl azide and two different cycloalkynes. a,b, TS structures for the reaction between methyl azide (Az) and the cycloalkynes cycloheptyne (7yne; **a**) and cyclononyne (9yne; **b**). In the diagram, cycloheptyne is shown in black and cyclononyne is shown in red. **c–f**, The resulting ASA (**c**) and EDA (**d**) graphs for the cycloaddition reactions, and the MO diagrams of the HOMO–FMO$_{yne}$ interaction (**e**) and the HOMO$_{yne}$–LUMO$_{Az}$ interaction (**f**) for the cycloaddition reactions. All data computed at M06-2X/TZ2P//M06-2X/6-31+G(d). r, distance; S, orbital overlap. Image adapted from ref. [44] under a Creative Commons Attribution 4.0 license (https://creativecommons.org/licenses/by/4.0/legalcode).

Methods 13), which results in the approximate TS structure depicted in Fig. 4a. This approximate TS geometry serves as a good starting geometry for the TS search (TSRC) calculation (Steps 12–14) (Supplementary Methods 14). When all stationary points have been calculated (R, RC, TS, and PC), one can make a relative energy diagram, with corresponding molecular geometries, as shown in Fig. 4b (see Supplementary Data 1 for Cartesian coordinates).

To obtain the curvature of the PES, one should run an IRC calculation (Steps 15–20). In this example, the IRC calculation is performed in parallel, meaning a backward IRC calculation (Supplementary Methods 15) and a forward IRC calculation (Supplementary Methods 16) are calculated at the same time. The IRC calculations are complete when the statement 'finishing IRC path Forward/Backward' is shown in the output file (Fig. 4c). Finally, the PES is dissected into its different energy components using PyFrag 2019 (Steps 21–27) (Supplementary Methods 17). These energy terms are projected onto the stretch of the C–Cl bond (Fig. 4d,e) (Steps 28–30).

In our second example, the ASM–EDA procedure is applied to study the SPAAC reaction between a 1,3-dipole, methyl azide (Az), and two cycloalkynes, cycloheptyne (7yne) and cyclononyne (9yne) (Fig. 5a, b; see Supplementary Data 2 for Cartesian coordinates)[44]. The energy terms from the ASM–EDA analysis are projected on the average distance of the two newly forming C•••N bonds, which has been shown to be a critical reaction coordinate for this class of reaction because it is directly related to the progress of the reaction.

The resulting activation strain diagram shows that the cycloaddition barrier is higher for 9yne than for 7yne (Fig. 5c). The interaction energy, ΔE_{int}, is entirely responsible for the order in reaction barriers, as the strain energies, ΔE_{strain}, are superimposed. This prompted us to further analyze the ΔE_{int} using the energy decomposition method. This analysis revealed that the trends in the Pauli repulsion, ΔE_{Pauli}, are more or less offset by the trends in the electrostatic interaction, ΔV_{elstat} (Fig. 5d). The orbital interactions, ΔE_{oi}, on the other hand, play a dominant role in determining the trends in interaction energy and, thus, the height of the reaction barriers.

The trend in ΔE_{oi} was rationalized by performing a Kohn–Sham MO analysis, which revealed that a greater pre-distortion, or bending of the alkyne, from 7yne to 9yne, leads to both a smaller FMO gap and an enhanced orbital overlap with the Az. Both the normal electron demand (Fig. 5e), between the $HOMO_{Az}$ (highest occupied MO of Az) and the unoccupied FMO_{yne} (unoccupied π^* MO of yne) and the inverse electron demand (Fig. 5f), between the $HOMO_{yne}$ and $LUMO_{Az}$ (highest occupied MO of yne and lowest unoccupied MO of Az, respectively), are smaller and thus more favorable for the cycloaddition between Az and 7yne. The larger degree of pre-distortion for 7yne results in a more destabilized $HOMO_{7yne}$ and more stabilized $LUMO_{7yne}$ as compared with that of 9yne, which, consequently, leads to a smaller FMO gap for the former. In addition, the greater degree of pre-distortion effectively polarizes the $HOMO_{7yne}$ and $LUMO_{7yne}$ lobes of 7yne toward the incoming Az, leading to an enhanced orbital overlap as compared with 9yne. These combined effects lead to enhanced orbital interactions that become more stabilizing as the ring size of the alkyne is decreased. Thus, we have shown how a combined ASM–EDA approach can be used to determine the physical mechanism underlining a particular chemical transformation.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support this study are available from the corresponding author upon reasonable request.

Code availability

The open-source PyFrag 2019 code that was used to generate all the data shown in the protocol can be found at https://github.com/sunxb05/PyFrag.

References

1. Albright, T. A., Burdett, J. K. & Wangbo, W. H. *Orbital Interactions in Chemistry* 2nd edn (Wiley, New York, 2013).
2. Piela, L. *Ideas of Quantum Chemistry* 2nd edn (Elsevier, Amsterdam, 2013).
3. Becke, A. D. Perspective: fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* **140**, 18A301 (2014).
4. Goerigk, L. & Grimme, S. Efficient and accurate double-hybrid-meta-GGA density functionals—evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **7**, 291–309 (2011).
5. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermo-chemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and

systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).

6. Perdew, J. P. & Constantin, L. A. Laplacian-level density functionals for the kinetic energy density and exchange-correlation energy. *Phys. Rev. B*. **75**, 155109 (2007).

7. Becke, A. D. Density functionals for static, dynamical, and strong correlation. *J. Chem. Phys.* **138**, 074109 (2013).

8. Zhao, Y., Schultz, N. E. & Truhlar, D. G. Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *J. Chem. Phys.* **123**, 16110 (2005).

9. Zhao, Y. & Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **41**, 157–167 (2008).

10. Hoffmann, R. Building bridges between inorganic and organic chemistry. *Angew. Chem. Int. Ed. Engl.* **21**, 711–724 (1982). *Angew. Chem.* **94**, 725–739 (1982)..

11. Fukui, K. The role of frontier molecular orbitals in chemical reactions. *Angew. Chem. Int. Ed. Engl.* **21**, 801–809 (1982). *Angew. Chem.* **94**, 852–861 (1982).

12. Marcus, R. A. On the theory of oxidation-reduction reactions involving electron transfer. I. *J. Chem. Phys.* **24**, 966–978 (1956).

13. Marcus, R. A. Electrostatic free energy and other properties of states having nonequilibrium polarization. I. *J. Chem. Phys.* **24**, 979–989 (1956).

14. Pross, A. & Shaik, S. S. A qualitative valence-bond approach to organic reactivity. *Acc. Chem. Res.* **16**, 363–370 (1983).

15. Usharani, D., Janardanan, D., Li, C. & Shaik, S. A theory for bioinorganic chemical reactivity of oxometal complexes and analogous oxidants: the exchange and orbital-selection rules. *Acc. Chem. Res.* **46**, 471–482 (2013).

16. Bickelhaupt, F. M. Understanding reactivity with Kohn-Sham molecular orbital theory: E2-S_N2 mechanistic spectrum and other concepts. *J. Comput. Chem.* **20**, 114–128 (1999).

17. Bickelhaupt, F. M. & Houk, K. N. Analyzing reaction rates with the distortion/interaction-activation strain model. *Angew. Chem. Int. Ed.* **56**, 10070–10086 (2017). *Angew. Chem.* **129**, 10204–10221 (2017).

18. Wolters, L. P. & Bickelhaupt, F. M. The activation strain model and molecular orbital theory. *WIREs Comput. Mol. Sci.* **5**, 324–343 (2015).

19. Fernández, I. & Bickelhaupt, F. M. The activation strain model and molecular orbital theory: understanding and designing chemical reactions. *Chem. Soc. Rev.* **43**, 4953–4967 (2014).

20. ADF, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com

21. te Velde, G. et al. Chemistry with ADF. *J. Comput. Chem.* **22**, 931–967 (2001).

22. Bickelhaupt, F. M. & Baerends, E. J. Kohn-Sham density functional theory: predicting and understanding chemistry. In *Reviews in Computational Chemistry*, Vol. 15 (eds Lipkowitz, K. B. & Boyd, D. B) 1-86. (Wiley-Blackwell, Hoboken, NJ, 2000)

23. Sun, X., Soini, T. M., Poater, J., Hamlin, T. A. & Bickelhaupt, F. M. Pyfrag 2019–automating the exploration and analysis of reaction mechanisms. *J. Comput. Chem.* **40**, 2227–2233 (2019).

24. Sun, X., Soini, T. M., Poater, J., Hamlin, T. A. & Bickelhaupt F. M. PyFrag 2019 [tutorial], (https://pyfragdocument.readthedocs.io/en/latest/includeme.html).

25. Fonseca Guerra, C., Handgraaf, J.-W., Baerends, E. J. & Bickelhaupt, F. M. Voronoi deformation density (VDD) charges. Assessment of the Mulliken, Bader, Hirshfeld, Weinhold and VDD methods for charge analysis. *J. Comput. Chem.* **25**, 189–210 (2004).

26. Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **44**, 129–138 (1977).

27. Swart, M., van Duijnen, P. Th & Snijders, J. G. A charge analysis derived from an atomic multipole expansion. *J. Comput. Chem.* **22**, 79–88 (2001).

28. Cabrera-Trujillo, J. J. & Fernández, I. Influence of the Lewis acid/base pairs on the reactivity of geminal E-CH_2–E′ frustrated Lewis pairs. *Chem. Eur. J.* **24**, 17823–17831 (2018).

29. García-Rodeja, Y., Solà, M. & Fernández, I. Understanding the reactivity of planar polycyclic aromatic hydrocarbons: towards the graphene limit. *Chem. Eur. J.* **22**, 10572–10580 (2016).

30. García-Rodeja, Y., Solà, M. & Fernández, I. Influence of the charge on the reactivity of azafullerenes. *Phys. Chem. Chem. Phys.* **20**, 28011–28018 (2018).

31. García-Rodeja, Y. & Fernández, I. Factors controlling the reactivity of strained-alkyne embedded cyclo-paraphenylenes. *J. Org. Chem.* **84**, 4330–4337 (2019).

32. Larrañaga, O. & de Cózar, A. Effect of an α-methyl substituent on the dienophile on Diels-Alder endo:exo selectivity. *ChemistryOpen* **8**, 49–57 (2019).

33. Jin, R., Liu, S. & Lan, Y. Distortion–interaction analysis along the reaction pathway to reveal the reactivity of the Alder-ene reaction of enes. *RSC Adv.* **5**, 61426–61435 (2015).

34. Liu, S., Lei, Y., Qi, X. & Lan, Y. Reactivity for the Diels–Alder reaction of cumulenes: a distortion-interaction analysis along the reaction pathway. *J. Phys. Chem. A* **118**, 2638–2645 (2014).

35. Galabov, B., Koleva, G., Schaefer, H. F. III & Allen, W. D. Nucleophilic influences and origin of the S_N2 allylic effect. *Chem. Eur. J.* **24**, 11637–11648 (2018).

36. Fell, J. S., Martin, B. N. & Houk, K. N. Origins of the unfavorable activation and reaction energies of 1-azadiene heterocycles compared to 2-azadiene heterocycles in Diels-Alder reactions. *J. Org. Chem.* **82**, 1912–1919 (2017).

37. Champagne, P. A. & Houk, K. N. Influence of endo- and exocyclic heteroatoms on stabilities and 1,3-dipolar cycloaddition reactivities of mesoionic azomethine ylides and imines. *J. Org. Chem.* **82**, 10980–10988 (2017).

38. Svatunek, D. & Houk, K. N. autoDIAS: a Python tool for an automated distortion/interaction activation strain analysis. *J. Comput. Chem.* **40**, 2509–2515 (2019).

39. Laloo, J. Z. A., Savoo, N., Laloo, N., Rhyman, L. & Ramasami, P. ExcelAutomat 1.3: fragment analysis based on the distortion/interaction-activation strain model. *J. Comput. Chem.* **40**, 619–624 (2019).

40. Liu, Y., Su, B., Dong, W., Li, Z. H. & Wang, H. Structural characterization of a boron (III) η2-σ-silane-complex. *J. Am. Chem. Soc.* **141**, 8358–8363 (2019).

41. Sun, X., Rocha, M. V. J., Hamlin, T. A., Poater, J. & Bickelhaupt, F. M. Understanding the differences between iron and palladium in cross-coupling reactions. *Phys. Chem. Chem. Phys.* **21**, 9651–9664 (2019).

42. Hong, X., Chan, K., Tsai, C. & Nørskov, J. K. How doped MoS_2 breaks transition-metal scaling relations for CO_2 electrochemical reduction. *ACS Catal.* **6**, 4428–4437 (2016).

43. Vermeeren, P., Sun, X. & Bickelhaupt, F. M. Arylic C–X bond activation by palladium catalysts: activation-strain analyses of reactivity trends. *Sci. Rep.* **8**, 10729 (2018).

44. Hamlin, T. A., Levandowski, B. J., Narsaria, A. K., Houk, K. N. & Bickelhaupt, F. M. Structural distortion of cycloalkynes influences cycloaddition rates both by strain and interaction energies. *Chem. Eur. J.* **25**, 6342–6348 (2019).

45. Hamlin, T. A., Fernández, I. & Bickelhaupt, F. M. How dihalogens catalyze Michael addition reactions. *Angew. Chem. Int. Ed.* **58**, 8922–8926 (2019). *Angew. Chem.* **131**, 9015-9020 (2019).

46. van der Lubbe, S. C. C. & Fonseca Guerra, C. Hydrogen-bond strength of CC and GG pairs determined by steric repulsion: electrostatics and charge transfer overruled. *Chem. Eur. J.* **23**, 10249–10253 (2017).

47. van der Lubbe, S. C. C., Zaccaria, F., Sun, X. & Fonseca Guerra, C. Secondary electrostatic interaction model revised: prediction comes mainly from measuring charge accumulation in hydrogen-bonded monomers. *J. Am. Chem. Soc.* **141**, 4878–4885 (2019).

48. Hamlin, T. A., van Beek, B., Wolters, L. P. & Bickelhaupt, F. M. Nucleophilic substitution in solution: activation strain analysis of weak and strong solvent effects. *Chem. Eur. J.* **24**, 5927–5938 (2018).

49. Laloo, J. Z. A., Rhyman, L., Ramasami, P., Bickelhaupt, F. M. & de Cózar, A. Ion-pair S_N2 substitution: activation strain analyses of counter-ion and solvent effect. *Chem. Eur. J.* **22**, 4431–4439 (2016).

50. Zaccaria, F., Paragi, G. & Fonseca Guerra, C. The role of alkali metal cations in the stabilization of guanine quadruplexes: why K^+ is the best. *Phys. Chem. Chem. Phys.* **18**, 20895–20904 (2016).

51. Zaccaria, F. & Fonseca Guerra, C. RNA versus DNA G-quadruplex: the origin of increased stability. *Chem. Eur. J.* **24**, 16315–16322 (2018).

52. Pollice, R., Bot, M., Kobylianskii, I. J., Shenderovich, I. & Chen, P. Attenuation of London dispersion in dichloromethane solutions. *J. Am. Chem. Soc.* **139**, 13126–13140 (2017).

53. Ess, D. H. & Houk, H. K. Distortion/interaction energy control of 1,3-dipolar cycloaddition reactivity. *J. Am. Chem. Soc.* **129**, 10646–10647 (2007).

54. Ess, D. H. & Houk, H. K. Theory of 1,3-dipolar cycloadditions: distortion/interaction and frontier molecular orbital models. *J. Am. Chem. Soc.* **130**, 10187–10198 (2008).

55. Hamlin, T. A. et al. Elucidating the trends in reactivity of aza-1,3-dipolar cycloadditions. *Eur. J. Org. Chem.* **2019**, 378–386 (2019).

56. de Jong, G. Th, Solà, M., Visscher, L. & Bickelhaupt, F. M. Ab initio benchmark study for the oxidative addition of CH_4 to Pd: importance of basis-set flexibility and polarization. *J. Chem. Phys.* **121**, 9982–9992 (2004).

57. de Jong, G. Th, Geerke, D. P., Diefenbach, A. & Bickelhaupt, F. M. DFT benchmark study for the oxidative addition of CH4 to Pd. Performance of various density functionals. *Chem. Phys.* **313**, 261–270 (2005).

58. de Jong, G. Th, Geerke, D. P., Diefenbach, A., Solà, M. & Bickelhaupt, F. M. Oxidative addition of the ethane C–C bond to Pd. An ab initio benchmark and DFT validation study. *J. Comput. Chem.* **26**, 1006–1020 (2005).

59. de Jong, G. Th & Bickelhaupt, F. M. Oxidative addition of fluoromethane C–F bond to Pd. An ab initio benchmark and DFT validation study. *J. Phys. Chem. A.* **109**, 9685–9699 (2005).

60. de Jong, G. Th & Bickelhaupt, F. M. Oxidative addition of the chloromethane C–Cl bond to Pd, an ab initio benchmark and DFT validation. *J. Chem. Theory Comput.* **2**, 322–335 (2006).

61. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).

62. Johnson, E. R. & Becke, A. D. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **123**, 024101 (2005).

63. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).

64. Bento, A. P., Solà, M. & Bickelhaupt, F. M. Ab initio and DFT benchmark study for nucleophilic substitution at carbon (S_N2@C) and silicon (S_N2@Si). *J. Comput. Chem.* **26**, 1497–1504 (2005).

65. Bento, A. P., Solà, M. & Bickelhaupt, F. M. E2 and S_N2 reactions of X^- + CH_3CH_2X (X = F, Cl). An ab initio and DFT benchmark study. *J. Chem. Theory Comput.* **4**, 929–940 (2008).

66. Pieniazek, S. N., Clemente, F. R. & Houk, K. N. Sources of error in DFT computations of C–C bond formation thermochemistries: π→σ transformations and error cancellation by DFT methods. *Angew. Chem. Int. Ed.* **47**, 7746–7749 (2008). *Angew. Chem.* **120**, 7860-7863 (2008).

67. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-019-0265-0.

Correspondence and requests for materials should be addressed to FM.B. or T.A.H.

Peer review information *Nature Protocols* thanks Xin Hong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Related links
Key references using this protocol
Vermeeren, P., Sun, X. & Bickelhaupt, F. M. *Sci. Rep.* **8**, 10729 (2018): https://doi.org/10.1038/s41598-018-28998-3
Sun, X., Soini, T. M., Poater, J., Hamlin, T. A. & Bickelhaupt, F. M. *J. Comput. Chem.* **40**, 2227–2233 (2019): https://doi.org/10.1002/jcc.25871

Key data used in this protocol
Vermeeren, P., Sun, X. & Bickelhaupt, F. M. *Sci. Rep.* **8**, 10729 (2018): https://doi.org/10.1038/s41598-018-28998-3
Hamlin, T. A., Levandowski, B. J., Narsaria, A. K., Houk, K. N. & F. Bickelhaupt, F. M. *Chem. Eur. J.* **25**, 6342–6348 (2019): https://doi.org/10.1002/chem.201900295

nature research

Corresponding author(s): F. Matthias Bickelhaupt and Trevor A. Hamlin

Last updated by author(s): Sep 20, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The open-source PyFrag 2019 code that was used to generate all of the data shown in the protocol, can be found using the provided link: https://github.com/sunxb05/PyFrag.
Data analysis	No software was used for a data analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	n/a
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	n/a
Research sample	n/a
Sampling strategy	n/a
Data collection	n/a
Timing	n/a
Data exclusions	n/a
Non-participation	n/a
Randomization	n/a

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	n/a
Research sample	n/a
Sampling strategy	n/a
Data collection	n/a
Timing and spatial scale	n/a
Data exclusions	n/a
Reproducibility	n/a
Randomization	n/a
Blinding	n/a

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging