Received: 27 October 2021 | Revised: 1 December 2021 | Accepted: 12 January 2022

DOI: 10.1002/qua.26892

RESEARCH ARTICLE

International Journal of Quantum Chemistry WILEY

A bonding evolution theory study of the reaction between methylidyne radical, CH($X^2\Pi$), and cyclopentadiene, C_5H_6

Juan Andrés[1] | Vicent S. Safont[1] | Mónica Oliva[1] | Kacee L. Caster[2] | Fabien Goulay[2]

[1]Department of Physical and Analytical Chemistry, Universitat Jaume I, Castelló de la Plana, Spain

[2]Department of Chemistry, West Virginia University, Morgantown, West Virginia, USA

Correspondence
Vicent S. Safont, Department of Physical and Analytical Chemistry, Universitat Jaume I, Avda. Sos Baynat s/n, 12071, Castelló de la Plana, Spain.
Email: safont@uji.es

Funding information
Generalitat Valenciana, Grant/Award Number: AICO 2020/329; Ministerio de Ciencia e Innovación, Grant/Award Number: PGC2018-094417-B-I00; National Science Foundation, Grant/Award Number: CHE-1764178; Universitat Jaume I, Grant/Award Number: UJI-B2019-30

Abstract

In the present work, bonding evolution theory (BET) is applied to gain insight about the complex reaction between methylidyne radical, CH ($X^2\Pi$) and cyclopentadiene, C_5H_6. The novelty of this work is that all reaction pathways take place in the doublet electronic state and an unpaired electron is always present. Therefore, taking the aforementioned reaction as explicative example, we have shown how to apply the BET tool to these kinds of open-shell systems, by splitting the wavefunctions into the corresponding alpha and beta parts. As an added value, we have included a point-by-point description of the algorithm we use to make it available for the readers. Hence, a complete analysis of bond breaking/forming and charge redistribution along the multi-channels connecting reactants to products via the transition states and intermediates is presented. We show how the BET brings about the representation of the electronic flow in complex molecular rearrangements like the one herein studied, yielding a transparent rationalization based on the electron density redistribution. The present study allows us to conclude that along the different processes giving rise to the benzene product, the breaking of a C—C sigma bond initiates the electronic rearrangement in two cases, but not in the third one. The last step in these processes can be described as an initial weakening of the C—H bond with a quasi-hydride formation and a final retro-transfer of electrons from the quasi-hydride to the C—H bond. On the other hand, in the way to the fulvene product, the breaking of the C—C sigma bond takes place after previous electronic redistribution. Neither the last step of the fulvene formation process nor the interesting H transfer described in the second one, can be explained without the wavefunction splitting technique herein detailed and exemplified.

KEYWORDS
BET, bonding evolution theory for open-shell systems, curly arrows, cyclopentadiene, methylidyne radical, molecular mechanism

1 | INTRODUCTION

Understanding the nature of chemical reaction mechanisms at atomic level has become one of chemistry's central research topics [1]. It consists not only of characterizing the main stationary points on a potential energy surface, but also in obtaining a detailed knowledge of how and where bond formation and bond cleavage processes take place. Chemistry, as many other disciplines in science, uses models, concepts, and languages based on physical grounds, which might constitute a more suitable representation. Curly arrows formalism is a paradigmatic example of a simple model to represent reaction mechanisms that facilitate a reliable prediction and rationalization of reaction mechanisms. The curly arrow represents the movement of electrons along a given chemical rearrangement, with that movement signifying the breaking and formation of bonds, providing an explanatory reaction mechanism [2].

In theoretical and computational chemistry, the study of the electronic structure and the chemical reactivity can be based on the electron density, $\rho(r)$. In contrast to the electronic wave-function, $\rho(r)$ is a physical observable, and therefore, represents a well-defined property for analysis and is an experimentally accessible scalar field and a local function defined within the exact many-body theory. In this context, quantum chemical topology (QCT), named by Popelier [3] is an established approach for the investigation of chemical bonding. It is capable to offer a partition of a chemical system into atomic regions and shows how to extract the information concerning their interaction.

The principal theoretical approach adopted by our research group while studying the nature of reaction mechanisms is the formulation and application of the bonding evolution theory (BET) [4]. The conceptual genesis of BET lies at the combination of the topological analysis of the electron localization function (ELF) [5] with the René Thom's catastrophe theory (CT) [6], and provides a powerful method that offers insight into the molecular mechanism of chemical rearrangements based on detailed physical grounds [7]. BET approach is capable to retrieve the valence-shell electron-pair repulsion (VSEPR) interpretation of chemical structure [8], and the classical curly arrows used to describe the rearrangements of chemical bonds. Furthermore, it provides detailed physical grounds for this type of representation [9] that seems to describe a more complete picture of the interplay between electron and nuclei rearrangements that characterize any chemical transformation [8–9, 10].

Within the BET methodology, the reaction mechanism becomes described in terms of chemical events or electronic transformations associated with the creation/annihilation of lone pairs and chemical bonding processes as the reaction evolves along a given pathway, enabling a practical representation of electron pair rearrangements occurring along the chemical transformation [7e]. The reaction pathway is decomposed in successive Structural Stability Domains (SSDs) within which the topological description of the electron density of the system does not vary. Each point between a SSD and the following is described as a catastrophe or turning point. The characterization of the successive SSDs and the corresponding turning points allows a deep analysis of electron density flows as a function of reaction progress, while chemical events such as breaking/forming bonds and/or rearrangements of pairs of electrons are monitored.

The CH radical was the first radical detected in the interstellar medium [11] and it plays a major role in combustion chemistry [12]. Its reaction with unsaturated hydrocarbons is believed to proceed through cycloaddition onto carbon double bonds or insertion into C—H or C—C single bonds. Although these mechanisms have been investigated both experimentally and theoretically [13], uncertainties remain on the competitiveness of the insertion mechanisms relatively to the cycloaddition mechanism.

Recently, Caster et al. [14] have investigated the reaction of the ground state methylidyne radical, CH ($X^2\Pi$) with cyclopentadiene, C_5H_6, in a quasi-static reaction cell at pressures ranging from 2.7 to 9.7 Torr and temperatures ranging from 298 to 450 K. Through ring expansion, the reaction is proposed as a novel route for benzene (C_6H_6) formation in combustion environments. To complement the experimental results and in order to clarify the nature of the corresponding reaction mechanism, involving complex reactive channels, such as ring-insertion and isomerization reactions, the authors performed first principle calculations. The combination of the experimentally determined rate constants and computed potential energy surfaces supports a fast, barrierless entrance channel that is characteristic of CH radical reactions and could potentially lead to the formation of benzene isomers. Despite the wealth of knowledge available discussing every aspect of this reaction, it is still widely studied because many crucial aspects remain unclear. In particular, essential features such as the evolution of bonding along the chemical transformation are not available.

We use BET to gain insight, proving its applicability to obtain a complete picture on the reactive pathways for the rationalization of non-intuitive complex reactivity patterns like the one herein studied, in which all paths take place in the doublet electronic state, because an unpaired electron is always present. Therefore, we have taken the reaction between methylidyne radical, CH ($X^2\Pi$) and cyclopentadiene, C_5H_6 as an explicative example to show how to apply the BET tool to these kinds of open-shell systems. A complete analysis of bond breaking/forming and charge redistribution along the multi-channels connecting reactants to products via the transition states and possible intermediates is presented, to shed light onto mechanistic features in these chemical reactions.

2 | COMPUTATIONAL METHODS

The geometry optimizations of the stationary points were performed by Caster et al. [14] using the Gaussian 09 program [15]. At the unrestricted B3LYP/CBSB7 level that selects the 6-311G(2d,d,p) basis set used by CBS-QB3 high accuracy energy method. The IRC paths have consequently

been now computed at the UB3LYP/6-311G(2d,d,p) theoretical level, using the second-order González-Schlegel integration method in order to obtain the energy profiles connecting each TS structure to the two associated minima of the corresponding path.

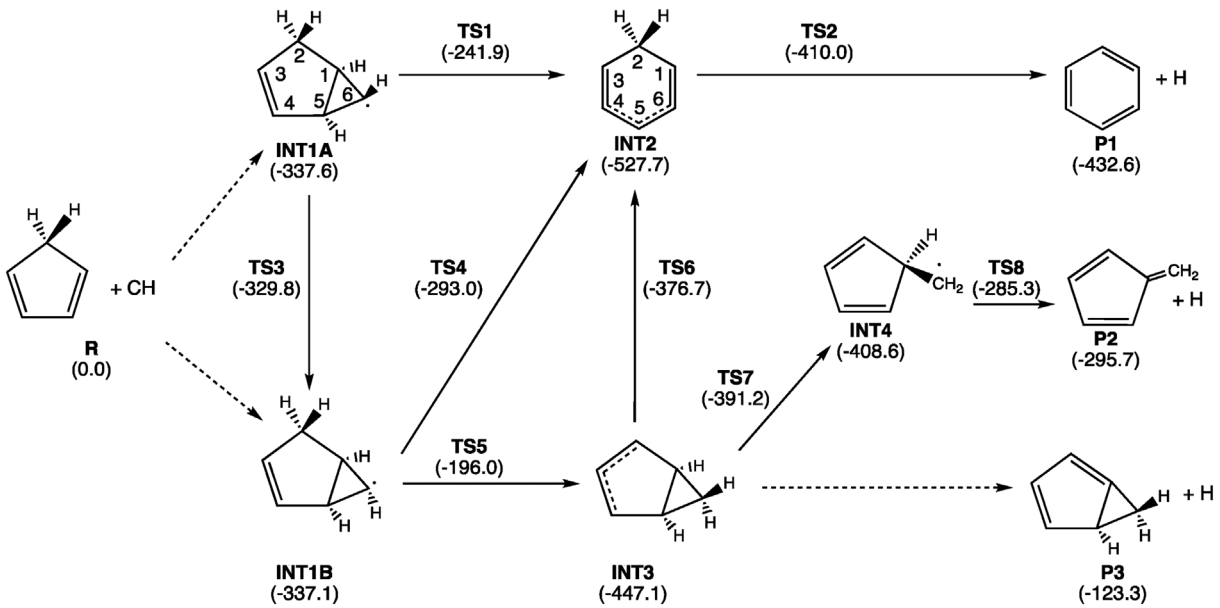
In the BET analysis, the wave function was obtained for each point of the IRCs and the ELF analysis was performed by means of the TopMod package [16] considering a cubical grid with a step-size smaller than 0.2 bohr. The detail of how we perform the BET analysis can be found elsewhere [7e, 7g, 9, 10c, 17]. As aforementioned, we have developed the BET tool for this open-shell system, and the wavefunctions are split into the corresponding α and β parts [7h], following the procedure detailed in the Appendix.

3 | RESULTS AND DISCUSSION

As Caster et al. concluded in their work [14], the CH cycloaddition, driven by strong attractive potential between the reactants, is likely to dominate over insertion or abstraction mechanisms at room temperature. In addition, recent product detection experiments coupled with computational studies [18] suggest that cycloaddition is likely to account for up to 90% of the overall mechanism. Therefore, we have chosen as a test processes on which to show the applicability of our methodology on open-shell systems, only the cycloaddition reaction between the CH ($X^2\Pi$) radical and cyclopentadiene. The mechanism is depicted in Scheme 1, with the C atoms numbering and the relative energies of the stationary points in parenthesis (kJ mol^{-1}). The names of the species participating in the mechanism are taken from the Caster work [14].

An analysis of the energy values showed that all pathways were found to be strongly exothermic, and all the stationary points found lied well below the energy of the reactants [14]. From the TSs geometries, we have now conducted the IRC calculations. In this way, we have linked the different intermediates through the corresponding TSs and have obtained the wavefunctions from each and every IRC point to conduct our topological study. In Scheme 1, the studied steps are indicated with solid-line arrows. The dashed-line arrows correspond to steps taking place without the participation of TSs: the CH radical interaction with cyclopentadiene renders either **INT1A** or **INT1B** in a barrierless way, while the **P3** formation from **INT3** consists in a H-loss (from C1) through a dissociation channel without TS. We have not included these steps in our topological study.

As can be seen in Scheme 1 the CH radical cycloaddition to cyclopentadiene can render two intermediates, **INT1A** or **INT1B**, depending on the CH substituent orientation. From **INT1A**, via the **TS1**, a ring opening can take place yielding **INT2**, which is much more stable due to resonance stabilization. Then, a C2 hydrogen elimination through **TS2** yields benzene (**P1**) and a hydrogen atom. The formation of benzene is the most favorable pathway. Intermediates **INT1A** and **INT1B** have very similar energies, and can be interconverted through **TS3**, with a very small energy barrier. From **INT1B**, two pathways are accessible: a ring opening process via **TS4**, or a H-transfer via **TS5**, rendering **INT2** or **INT3**, respectively. **INT3** and **INT2** can be connected via a ring opening path through **TS6**. Finally, from **INT3** two other pathways are also possible: another ring opening via **TS7** that gives **INT4**, which in turn can lose a H atom from C1 through **TS8** to render fulvene (**P2**) and a hydrogen atom, and the aforementioned H loss to obtain the bicyclic product (**P3**) and a hydrogen atom.



SCHEME 1 Complete mechanism for the cycloaddition reaction between CH ($X^2\Pi$) radical and cyclopentadiene

In what follows, we offer a detailed description of the different pathways from the topological point of view that the BET provides. As aforementioned, the wavefunctions are split into the corresponding α and β parts [7h] following the algorithm detailed in the Appendix, so that both topologies must be analyzed, and this fact makes the description of the process to be rather complex. We present our results in a simplified way without losing the key aspects of the bonding evolution along the different studied steps.

3.1 | INT1A **to** INT2 **through** TS1

The IRC traced from **TS1** includes 29 steps in the reverse direction to **INT1A** and 84 steps in the forward direction to **INT2**, see Scheme 2. Therefore, taking into account that we have split the wavefunctions into the corresponding alpha and beta parts, as aforementioned, the ELF function has been calculated for a total of 228 points, including the alpha and beta wavefunctions of the **TS1** itself. Up to 28 SSDs can be found along this process. In the 27 turning points between adjacent SSDs several kinds of catastrophes have been identified, and the whole set of data has been graphically included in the supporting information (see Figures S1 and S2).

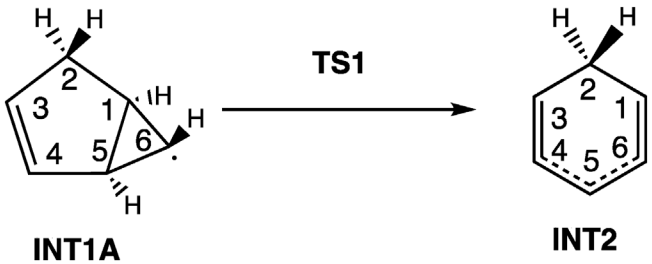
Among these turning points, it is worth noting that the more significative changes refer to the disappearance of the disynaptic V(C1,C5) basin that accounts for the breaking of this bond, the disappearance of the monosynaptic V(C6) basin due to the delocalization of the initial radical electron, and the appearance of the two V(C5) monosynaptic basins reflecting the accumulation of the unpaired electrons around this carbon atom in the final structure.

The V(C1,C5) basin disappears in the beta part of the electronic structure (at the 10th point of the reverse IRC, point −10, see Figure S2) before than in the alpha part (at the −8 point of the IRC, where the SSD-IV is reached). Therefore, the C1—C5 bond can be considered broken once SSD-IV is reached, because from this domain on, the V(C1,C5) basin cannot be found, irrespectively of considering the alpha or beta part of the wavefunction. This is the first significant chemical event taking place along the IRC pathway and at this point the C1—C5 distance is 1.999 Å. One of the important chemical contribution of the BET is to characterize when the single bonds are broken or formed, and in this case we can conclude that the C1—C5 bond breaks when its distance achieves that value.

In Figure 1, the evolution of the total population (i.e., the sum of the alpha and beta populations) of selected basins is depicted as a function of the IRC point number. As can be seen, we have grouped the populations of the basins V(C3,C4) and V(C3,C4)′, the populations of the basins V(C1,C6) and V(C1,C6)′, and the populations of the basins V(C5,C6) and V(C5,C6)′. These disynaptic basins account for double bonds and appear and disappear several times as the reaction takes place (see Figure S2). For instance, the disynaptic V(C3,C4)′ basin disappears in the alpha part of the wavefunction at SSD-VI, reappears at SSD-XIII, disappears again at SSD-XXVII, and finally appears at SSD-XXVIII. As for the beta part, this basin disappears at SSD-II, appears at SSD-XV, and finally disappears at SSD-XXII. In spite of this mess, the total electron population between the atoms C3 and C4 remains stable around 3.4 electrons along all the process: it diminishes more or less monotonically from 3.48 to 3.37, see Figure 1, blue line. Therefore, to avoid including non-significant data in the discussion, we report the evolution along the IRCs of the total electron population of the set of basins accounting for each initial, final or transient double bond.

The V(C1,C5) population (orange line) quickly decreases from the initial value of 1.65 at reactant to a final value of 0.16. This population has been transferred to the basins between C5 and C6, between C1 and C6, and to V(C4,C5) (pale green, red-brown, and violet lines in Figure 1, respectively).

The second significant chemical event taking place is the V(C6) monosynaptic basin disappearance, that takes place at the 25th point of the forward IRC, where the SSD-XIV is found. Its population (dark green line in Figure 1) is of 1.1 at reactant, diminishes along the initial domains until a minimum value is found by the 10th point of the forward IRC, concomitantly with an increase in the population of the basins between C5—C6 and between C1—C6, as can be seen in the Figure 1. After that the V(C6) population slightly increases, mainly at expense of the basins between C1 and C6, and diminishes again until a final value of 0.305. This residual population goes to the basins between C5 and C6 when V(C6) disappears.



SCHEME 2 **INT1A** to **INT2** step with atom numbering

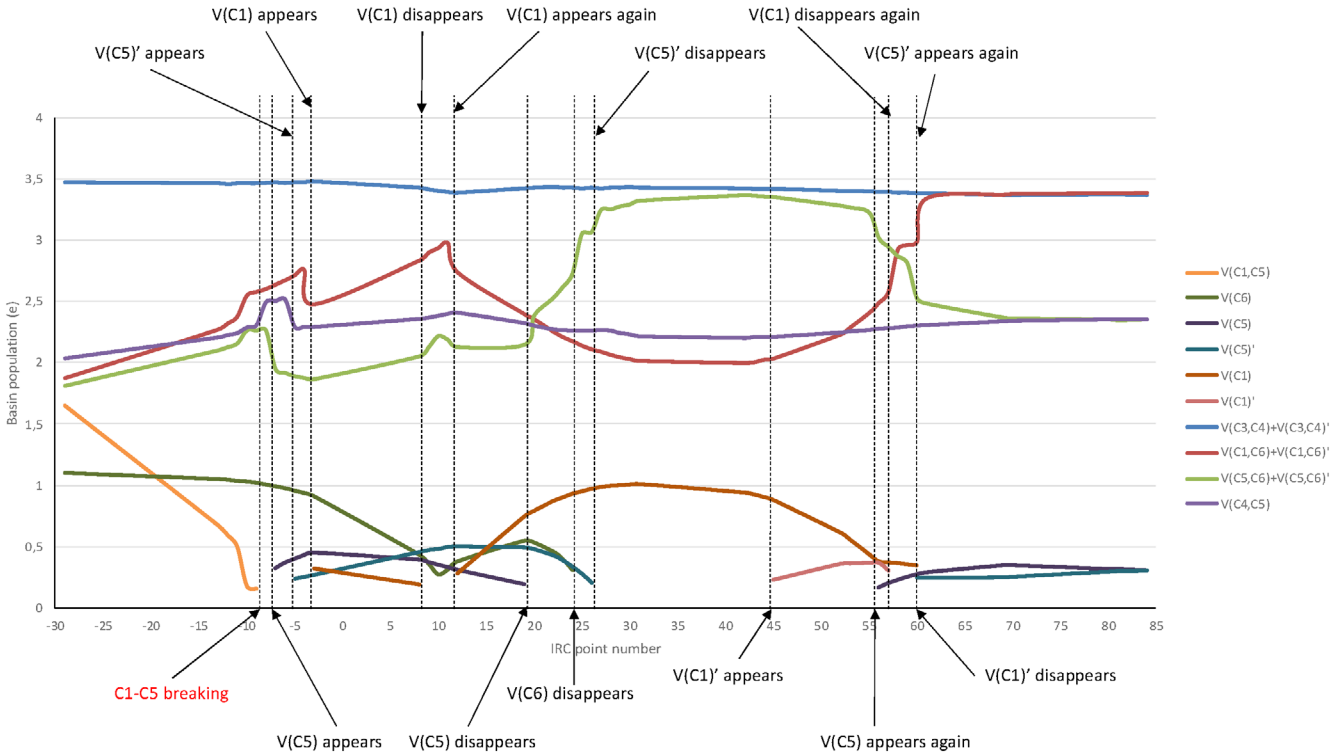


FIGURE 1 Basins population evolution along the IRC from **TS1**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The change reflecting the C1—C5 sigma bond disappearance is colored in red

The V(C5) and V(C5)′ monosynaptic basins appear at SSD-V and at SSD-VII, respectively, see Figure S2. However, thereafter these basins disappear at SSD-XII (20th point of the forward IRC), and at SSD-XVI (IRC point 27), respectively, due to internal electronic redistribution. And they reappear at SSD-XXIII (IRC point 56) and at SSD-XXV (IRC point 60), to reach the final electronic distribution. The initial population of V(C5) (dark violet line) is taken from the basins between C5 and C6 (pale green line in Figure 1), and slightly increases from 0.33 to 0.45, then decreases to 0.2 and it disappears, to reappear thereafter as explained. The second monosynaptic basin, V(C5)′, is located opposite to V(C5) with respect the C5 atom, see Figure S2, and takes its initial population (blue-gray line) from the V(C4,C5) basin (violet line). The population of this V(C5)′ basin also slightly increases from 0.24 to 0.5, decreases to 0.21 and then it disappears, appearing again as explained.

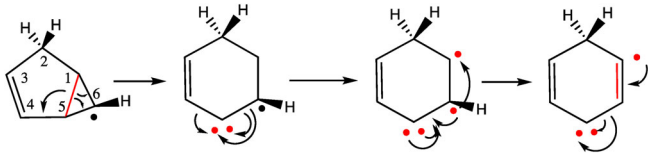
It is interesting to observe the evolution of the population of the multiple bonds along this reaction. As explained, the total electron population between C3 and C4 remains stable. The population between C1 and C6 (red-brown line in Figure 1) increases from 1.88 to 2.97, then decreases and increases following a symmetric tendency with respect to the population of the V(C1) monosynaptic basin (see below), and finally reaches the same value (3.38) than the population between C3 and C4. It is also interesting to observe that the population between C5 and C6 greatly increases from 1.81 to 3.36, and then decreases to 2.35. In fact, between these two atoms a transient second disynaptic basin V(C5,C6)′ appears along the SSD-XVII and SSD-XVIII, indicating that the C5—C6 bond can be considered as double along these two domains due to the transient population accumulation.

In addition, two transient monosynaptic basins are formed on C1: V(C1) and V(C1)′, located in opposite sides of C1 atom. They appear and disappear along the process and are not present neither at reactants nor at the products, see Figure 1.

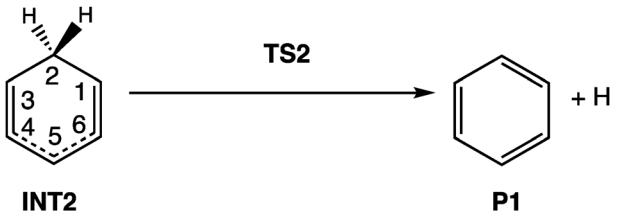
At **INT2** the electronic distribution reflects two nearly double bonds at C1—C6 and C3—C4, with populations of 3.38 and 3.37 respectively; two more-than-single bonds between C4—C5 and C5—C6, both with a population of 2.35 electrons; two single bonds between C1—C2 and C2—C3, both with a population of 1.98 electrons (not included in Figure 1); and two monosynaptic basins on C5, with small populations of 0.31 electrons, where the radical character is centered.

From this analysis it can be concluded that the C1—C5 breaking takes place well before the TS is reached, and is the key event leading, and causing, the electronic reorganization from **INT1A** to obtain **INT2**.

The main electron density flows can be schematically drawn as shown in Scheme 3 in four stages: (i) the early disappearance of the V(C1—C5) basin with electron density flowing to the V(C1—C6), V(C5—C6) and V(C4,C5) basins; (ii) the creation of the V(C5) and V(C5)′ basins from V(C5,C6) and V (C4,C5); (iii) the disappearance of V(C5) and V(C5)′ due to electron density flowing to V(C5,C6), as well as the disappearance of the initial V(C6) transferring its population to the transient V(C1) and also to V(C5,C6); (iv) the disappearance of the transient V(C1) giving its population to V(C1,C6), that becomes a double bond, concomitant with a population transfer from V(C5,C6) to the reappeared monosynaptic V(C5) and V(C5)′ basins. In Scheme 3, the disappearing or appearing basins are highlighted in red, the curved arrows stand for electron density flows and the fat dots stand for monosynaptic basins.



SCHEME 3 Main electron density flows for the **INT1A** to **INT2** step



SCHEME 4 **INT2** to **P1** step with atom numbering

3.2 | INT2 to P1 through TS2

The IRC from **TS2** includes 32 steps in the reverse direction to **INT2** and 22 steps in the forward direction to **P1**, see Scheme 4. Therefore, in this case, the ELF function has been calculated for a total of 110 points. This process has been found to be simpler than the preceding one, and only 5 SSDs have been found (see Figure S3). The more significant changes take place between SSD-II and SSD-III, where the two monosynaptic basins on C5 disappear, and at the turning point between SSD-IV and SSD-V, where the bond between the C2 and the leaving hydrogen atom breaks, as described by the definitive disappearance of the former V(C2,H) basin. At this point, the C2—H distance is 3.176 Å.

In Figure 2, the evolution of the population of selected basins along the IRC traced from **TS2** is depicted as a function of the IRC point. As can be seen in that figure, the disappearance of the monosynaptic basins on C5 takes place before the TS, and their populations were going from the beginning to the V(C4,C5) and the V(C5,C6) basins. The appearance of the V(H) basin accounting for the leaving of the hydrogen atom takes also place before the TS (at the IRC point −3, where the C2—H distance is 1.694 Å). However, the appearance of this V(H) basin does not imply the total disappearance of the hydrogenated V(C2,H) basin, because although an important amount of the population of V(C2,H) basin is transferred to V(H), it still remains a residual population in V(C2,H), indicating that the interaction between C2 and the leaving hydrogen is weakly maintained until the last point of the SSD-IV, as explained. The V(C2,H) basin was losing population from the beginning, as can be seen in the figure. This population was going to the V(C2,H′), V(C2,C3), and V(C1,C2) basins.

The V(H) population evolves from 0.91 when this basin appears taking its population from the V(C2,H) basin, then increasing to 1.45 and after that decreasing to a final value of 1.1. The initial increasing in the V(H) population is coupled with the V(C2,H′) basin depopulation. A final retrotransfer from V(H) to V(C2,H′) is sensed, and a single C2—H′ bond and a hydrogen atom are finally recovered. In other words, the electronic rearrangement accompanying the C2—H bond breaking can be described as an initial weakening of the C2—H bond coupled with a quasi-hydride formation, and a final returning of the hydride electronic excess to the C2—H′ bond.

Also from Figure 2, it is interesting to observe the lines representing the population evolution of the C—C bonds. The population of the former double bonds (between C1 and C6 and between C3 and C4) monotonically decreases from 3.38 to 2.77, while the population of the other C—C bonds increases: the V(C4,C5) and V(C5,C6) basins population evolves from *ca* 2.35 to 2.76, and the V(C1,C2) and V(C2,C3) population evolves from ca 1.97 to 2.76. The formation of benzene, with all C-C bonds equivalent, is paradigmatically reflected in Figure 2.

These electronic flows are indicated with curved arrows in Scheme 5, in which the appearing and disappearing basins are highlighted in red, and the final aromatic character of benzene is indicated with the round red arrow.

3.3 | INT1A to INT1B through TS3

The IRC from **TS3** includes 27 points in the reverse direction to **INT1A** and 24 points in the forward direction to **INT1B**, see Scheme 6, so that the ELF function has been calculated for a total of 104 points. The process is even simpler than the one just described, and only 3 SSDs have been found (see Figure S4).

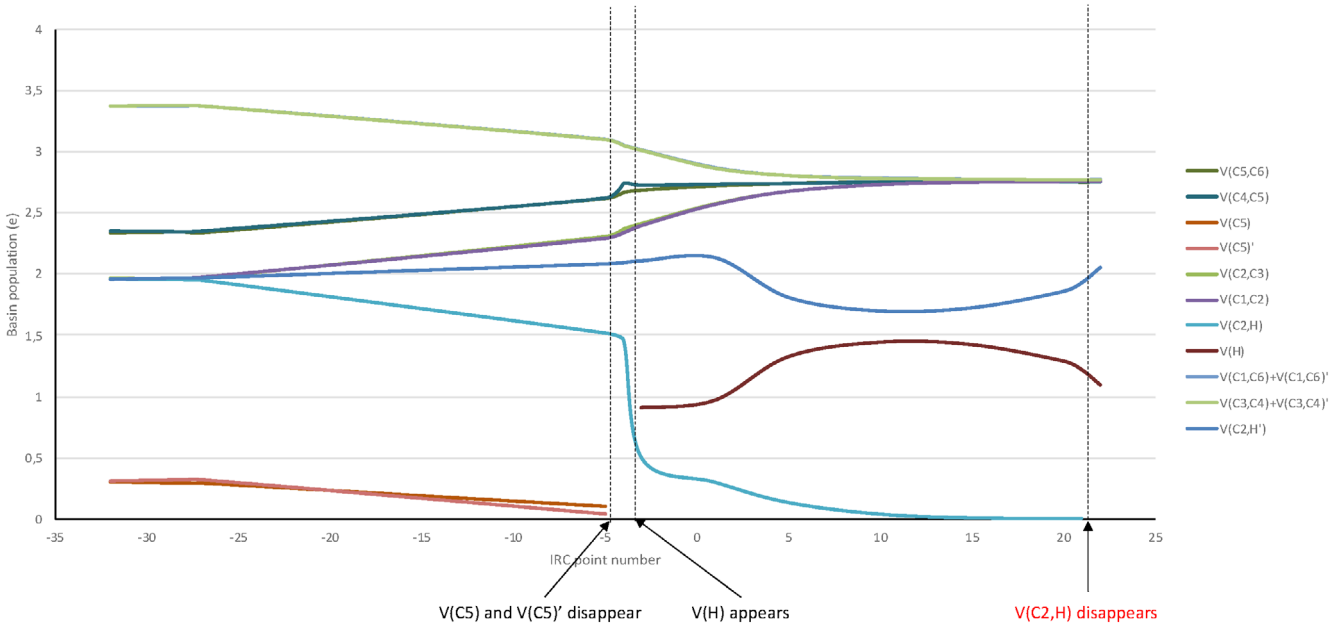
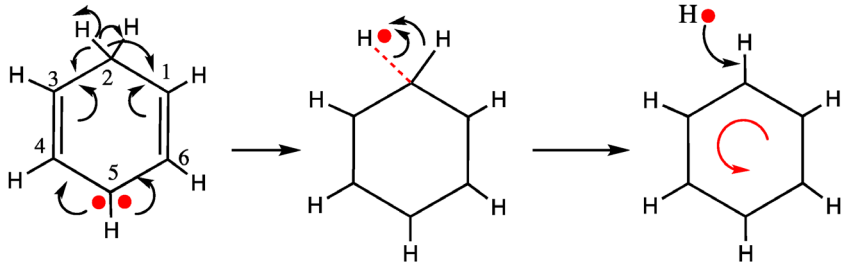
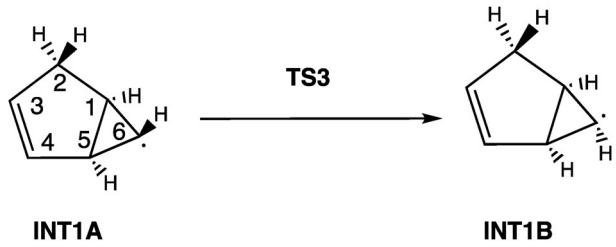


FIGURE 2 Basins population evolution along the IRC from **TS2**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The change reflecting the C2—H bond disappearance is colored in red



SCHEME 5 Main electron density flows for the **INT2** to **P1** step



SCHEME 6 **INT1A** to **INT1B** step with atom numbering

The two turning points are observed in the alpha part of the wavefunction and correspond to the appearance of a second monosynaptic basin on C6, V(C6)', and the disappearance of the former V(C6) monosynaptic basin. In Figure 3, the evolution of the population of selected basins are shown as a function of the IRC point.

The population needed for the V(C6)' creation comes from V(C6,H) that was taking population from V(C6) from the beginning of the process. Thereafter, the population of V(C6)' increases as the population of V(C6) decreases, and finally the population of the former V(C6) basin goes to V(C6,H) when V(C6) disappears. A final increase of the population of V(C6)' is sensed concomitantly with the decrease of the V(C6,H) population, to reach a situation analogous to the initial one, but with the positions of the H atom and the monosynaptic basin on C6 interchanged.

These electronic flows can also be schematically depicted with arrows, as shown in Scheme 7.

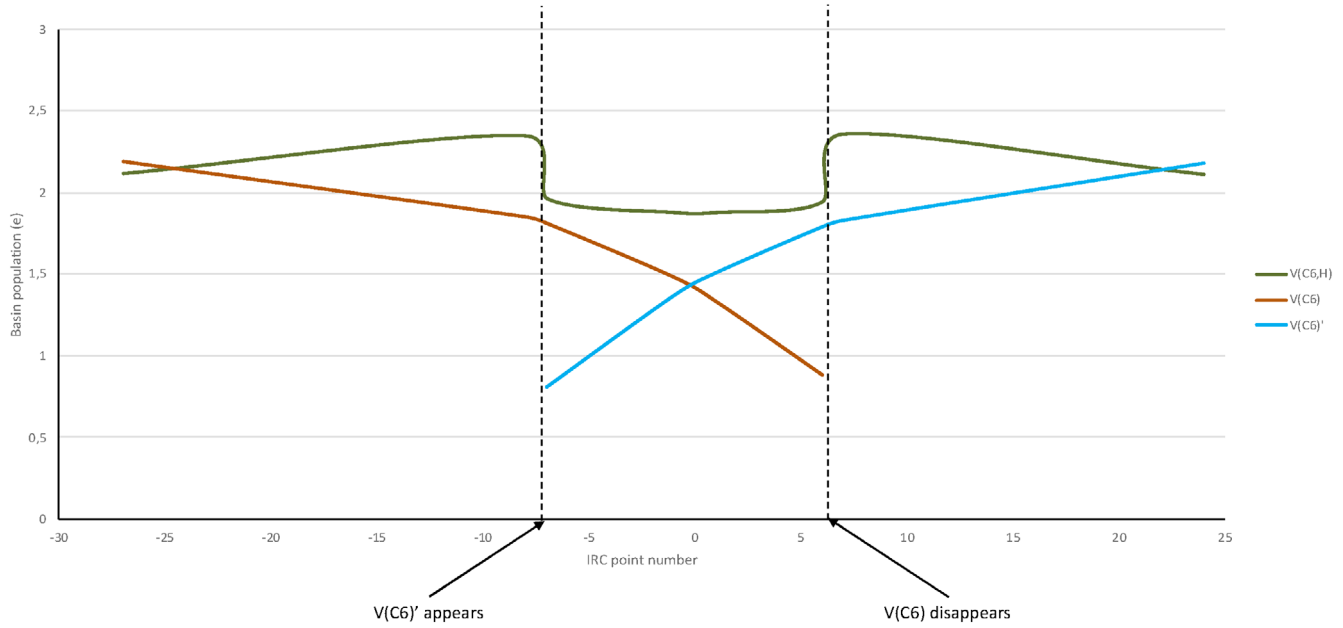
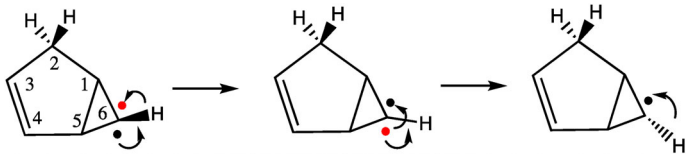
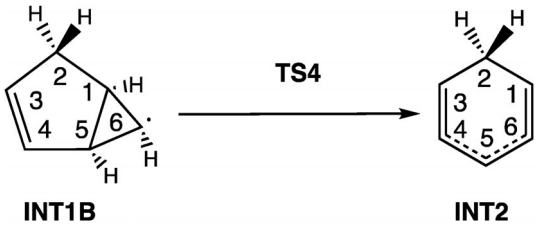


FIGURE 3 Basins population evolution along the IRC from **TS3**



SCHEME 7 Main electron density flows for the **INT1A** to **INT1B** step



SCHEME 8 **INT1B** to **INT2** step with atom numbering

3.4 | INT1B **to** INT2 **through** TS4

The IRC from **TS4** includes 34 points in the reverse direction to **INT1B** and 76 points in the forward direction to **INT2**, see Scheme 8, so that the ELF function has been calculated for a total of 222 points. The process should a priori be very similar to the one linking **INT1A** with **INT2** through **TS1**. However, in this case only 13 SSDs have been found (see Figures S5 and S6).

Among the 12 turning points, the more significative changes refer to the disappearance of the disynaptic V(C1,C5) basin, that takes definitively place in the alpha part of the electronic structure, at the turning point between SSD-IV and SSD-V, the disappearance of the monosynaptic V(C6) basin, that occurs at the frontier between SSD-VII and SSD-VIII, and the appearance of the two V(C5) and V(C5)'monosynaptic basins. These appear also at the turning point between SSD-IV and SSD-V, although V(C5)' disappears at the turning point between SSD-V and SSD-VI and appears again at the turning point between SSD-X and SSD-XI.

As can also be seen in Figure 4, the first significant chemical events taking place are the simultaneous breaking of the C1—C5 bond and the formation of the two monosynaptic V(C5) and V(C5)' basins. These events occur now at the 1st point of the forward IRC, when the C—C distance is 2.016 Å, a little bit larger than in the **TS1** case, in which the C-C distance was 1.999 Å. The V(C1,C5) basin population (olive green line) quickly decreases from the initial value of 1.61 at reactant to a final value of 0.3 at the TS of this process. As can be seen,

the population that V(C1,C5) loses goes initially to the basins between C1 and C6, between C5 and C6, and between C4 and C5 atoms (violet, pale blue and orange lines respectively). Finally, when this disynaptic basin disappears, its residual population is transferred to the newly created monosynaptic V(C5)' basin (blue line), while the V(C5,C6) basin depopulates because of the V(C5) appearance (sky blue line).

The V(C5) population increases initially at expense of the V(C6) basin (red line). After that, the V(C5) population monotonically decreases until a final value of 0.31, concomitantly with an increase in the population of the V(C5,C6) basin.

As for the V(C5)' monosynaptic basin, its first and very short appearance takes place along the two initial points of the forward IRC with a small population of ca 0.3 |e|. Then it disappears and its small population goes to the V(C4,C5) disynaptic basin. After that it appears again at the 47th point of the forward IRC with an initial population, taken from V(C4,C5) basin, of 0.17, that increases until a final value of 0.31, taking population mainly from the V(C5) basin.

The other significant chemical event taking place is the disappearance of the monosynaptic V(C6). Its population (red line) is of 1.1 at reactant and remains more or less constant along the initial domains. Then it diminishes until a value of 0.31 is found at the 9th point of the forward IRC. The population that this basin loses is gained by the disynaptic basins between C1 and C6 (violet line) and by V(C5), as explained.

It can be observed a transient appearance of a V(C3) basin (brilliant green line), with a small population of ca 0.3 that is taken from the basins between C3 and C4 (pale green line in Figure 4). It is also interesting to observe the final equilibration of the populations of the basins between C3 and C4 and the populations between C1 and C6, due to the final formation of two equivalent quasi-double bonds. Also, the populations of the basins between C4 and C5 and between C5 and C6 reach the same value (ca 2.4 |e|) at the end of the process because of the formation of two equivalent more-than-single bonds.

From this analysis it can be concluded that the C1—C5 breaking takes now place when the TS is reached, and is the key event leading, and causing, the electronic reorganization from **INT1B** to obtain **INT2**. The curly arrows describing the main electron flows can schematically be drawn as shown in Scheme 9: the simultaneous disappearance of C1—C5 bond and appearances of the monosynaptic basins on C5 with the indicated electron flows, the disappearance of V(C5)' populating V(C4,C5), the transient appearance of the V(C3) basin, the disappearance of V(C6) populating V(C5) and the basins between C1 and C6, and the final appearance of V(C5)' taking electrons from V(C4,C5) and V(C5).

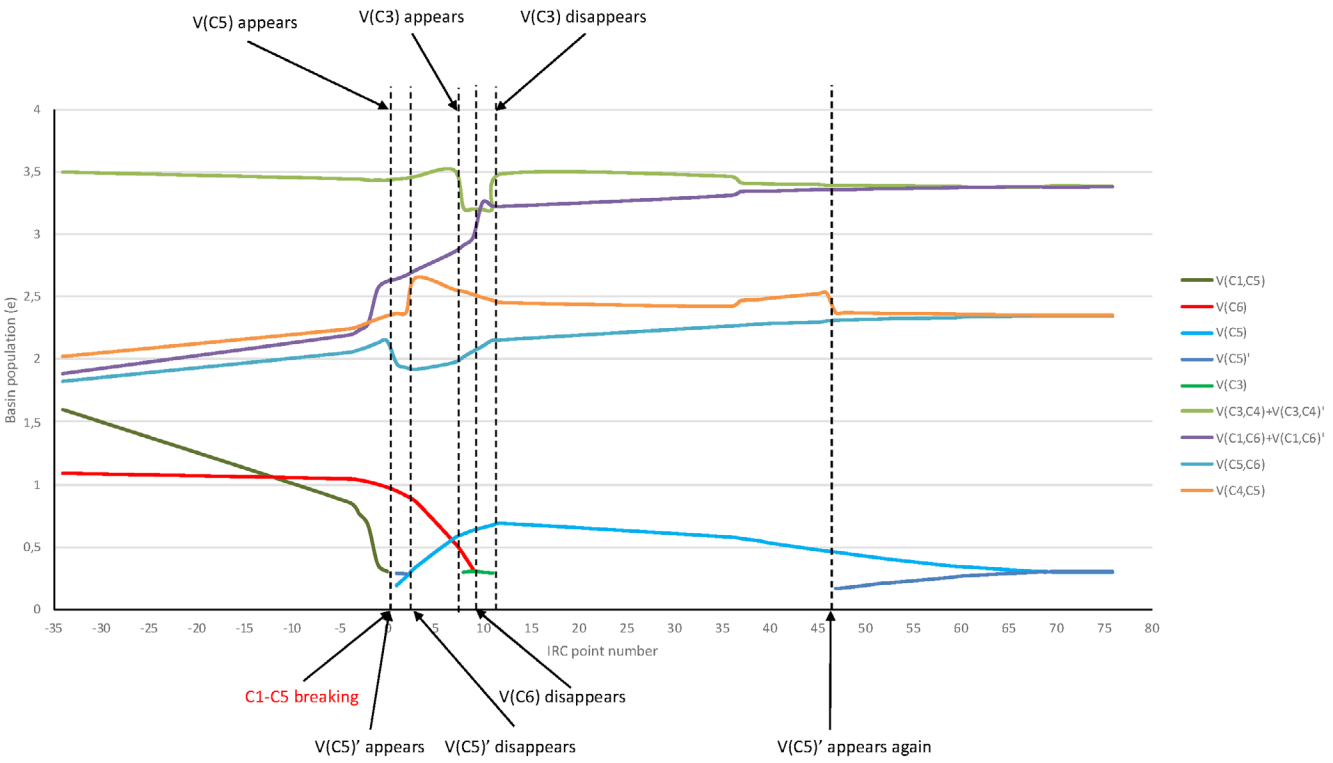


FIGURE 4 Basins population evolution along the IRC from **TS4**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The change reflecting the C1—C5 sigma bond disappearance is colored in red

3.5 | INT1B **to** INT3 **through** TS5

The IRC from **TS5** includes 79 points in the reverse direction to **INT1B** and 61 points in the forward direction to **INT3**, see Scheme 10, so that the ELF function has been calculated for a total of 282 points. Along the IRC, 12 SSDs have been found (see Figures S7 and S8).

Among the 11 turning points, the more significative changes refer to the H transfer from C2 to C6. As can be seen, this hydrogen transfer takes place along several topological changes: first, the early V(C2,H) basin is transformed into a V(C2,C6,H) basin in the beta part of the topology in the turning point between SSD-I and SSD-II, IRC point −3. At his point, the C2—H distance is 1.264 Å, while the C6—H distance is 1.596 Å. On the alpha part, on the contrary, the early V(C2,H) basin is transformed into a V(H) basin at SSD-IV. Finally, at SSD-VI, IRC point 6, both the V(H) basin in the alpha part and the V(C2,C6,H) basin on the beta part disappear giving birth to the V(C6,H) basin. At this point the distances C2—H and C6—H are, respectively, 1.509 and 1.358 Å.

The disappearance of the monosynaptic V(C6) basin as well as the appearances of the V(C4)' and V(C2)' monosynaptic basins are also important in the pathway to the final electronic distribution at **INT3.** The V(C6) disappearance takes place at the turning point between SSD-VI and SSD-VII, IRC point 7, while the appearances of the V(C4)' and V(C2)' monosynaptic basins occur when the SSD-VIII (IRC point 10) and SSD-IX (IRC point 15) domains, respectively, are reached.

As can be seen in Figure 5, the changes associated with the H transfer process take place in a very narrow zone of the IRC around the TS. There is a wide region before the TS in which no topological changes take place, and the system deforms, without changing its electronic structure, in order to approach the H being transferred to the C6 atom. Once the interaction between the C6 and the H is significant, the electronic distribution quickly changes, and the H transfer is completed. Then the system rearranges to form the final monosynaptic V(C2)' and V(C4)' basins, with transient V(C2) as well as V(C4) monosynaptic basins appearing and disappearing.

Regarding the motion of the electrons, the 0.86 |e| initial population of the V(C2,C6,H) basin is taken from the V(C2,H) disynaptic basin. Also, the V(C2) monosynaptic basin creation takes place at expense of the V(C2,H) basin, that is finally converted into the V(H) basin. The H transfer is completed with the creation of the V(C6,H) basin that takes its population from the disappearing V(C2,C6,H) and V(H) basins.

On the other hand, when the V(C6) monosynaptic basin disappears its population is transferred to the just created V(C6,H) basin, that achieves this way its final population of ca 2.1 |e|. The population needed for the creation of the V(C4)' monosynaptic basin (0.3 electrons) is taken from the disynaptic basin V(C3,C4), that is also the source for the electrons needed to create the transient V(C4) basin. When this transient basin disappears, its populations come back to the V(C3,C4) basin. Finally, as aforementioned the V(C2) is created taking its population from the former V(C2,H) basin, and when V(C2) disappears, its population goes to the V(C2,C3) disynaptic basin. And the V(C2)' monosynaptic basin takes its initial population from the V(C2,C3) basin, as can be seen in Figure 5.

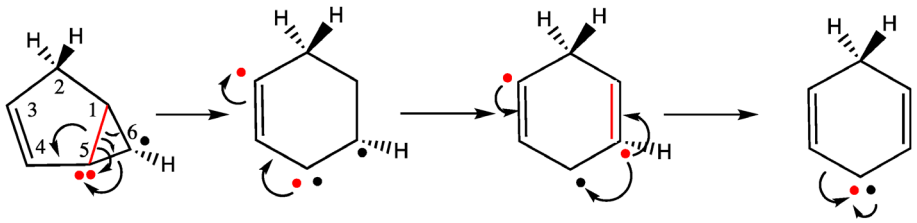
The curved arrows describing all these motions are shown in Scheme 11.

3.6 | INT3 **to** INT2 **through** TS6

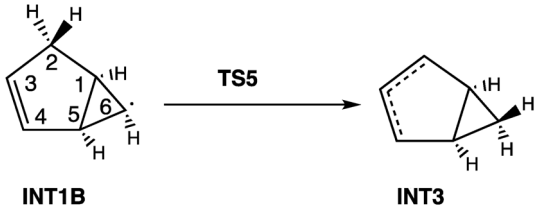
A word of caution must be said before explaining this step: although the final product of this step is **INT2**, we must be aware that the numbering of the atoms at **INT2** will now be different from the numbering shown in Scheme 1. To avoid confusions, the step is reproduced in the following Scheme 12 with the correct atoms numbering.

The IRC from **TS6** includes 39 points in the reverse direction to **INT3** and 119 points in the forward direction to **INT2**, so that the ELF function has been calculated for a total of 318 points. Along the IRC, 19 SSDs have been found (see Figures S9 and S10).

Among the 18 turning points, the more significant changes refer to the disappearance of the V(C1,C5) disynaptic basin accounting for the C1—C5 bond breaking, that takes place at the SSD-VII, IRC point −1, when the C1—C5 distance is 1.945 Å, and also to the disappearance of the initial V(C2)' and V(C4)' monosynaptic basins taking place at the SSD-IV (IRC point −13) and SSD-VIII (the TS), respectively, as well as to the final appearance of the V(C3) and V(C3)' monosynaptic basins, that can be observed at the SSD-XIV (IRC point 29) and SSD-XVI (IRC point 46), respectively.



SCHEME 9 Main electron density flows for the **INT1B** to **INT2** step



SCHEME 10 **INT1B** to **INT3** step with atom numbering

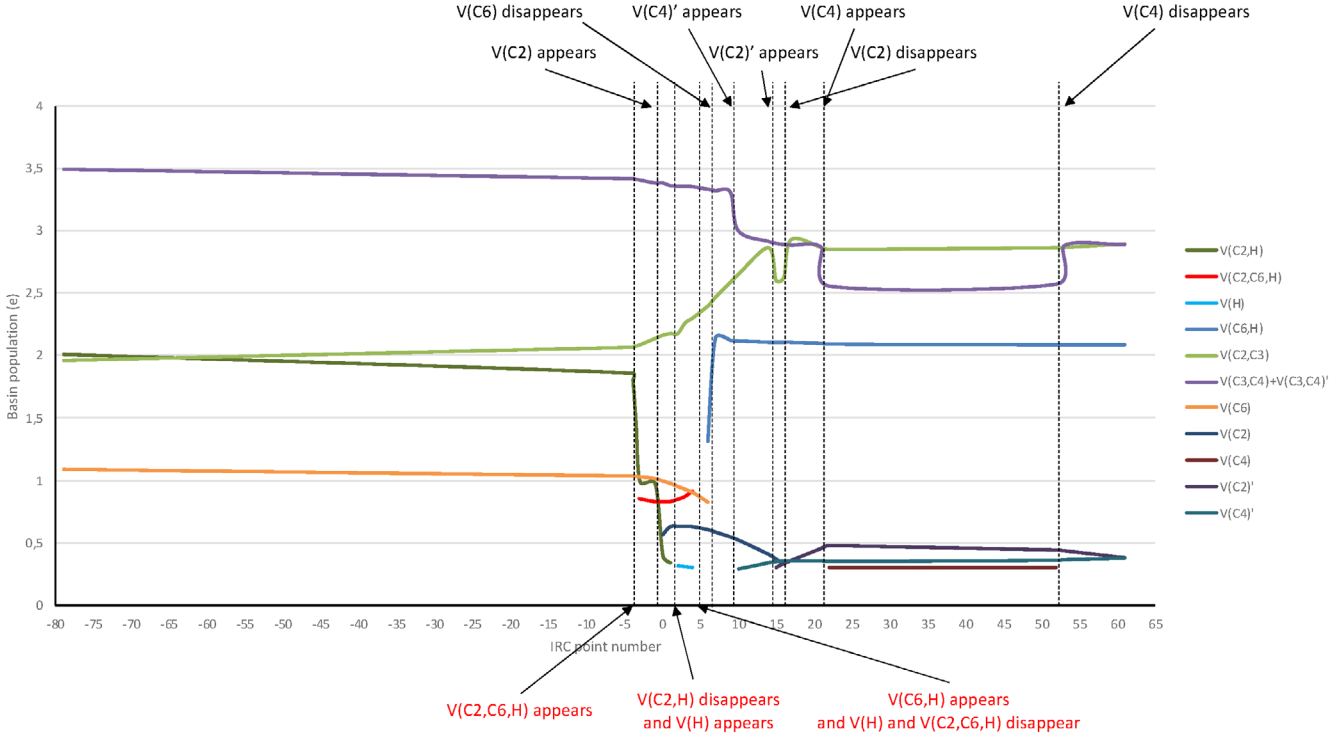
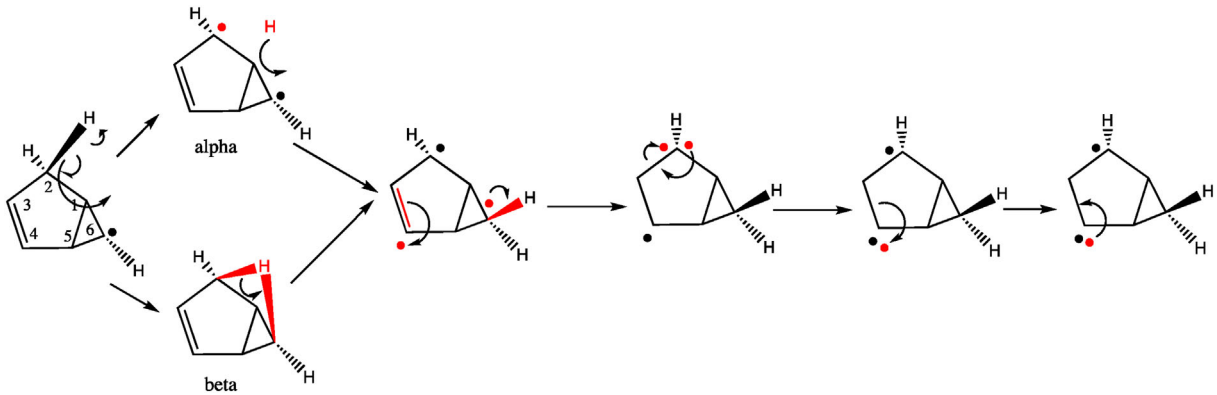
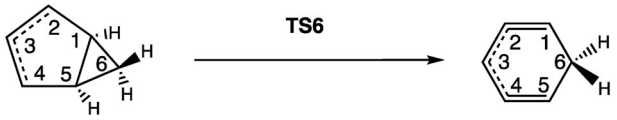


FIGURE 5 Basins population evolution along the IRC from **TS5**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The changes related with the C2—H breaking and C6—H forming bonds are colored in red



SCHEME 11 Main electron density flows for the **INT1B** to **INT3** step



SCHEME 12 **INT3** to **INT2** step with atom numbering

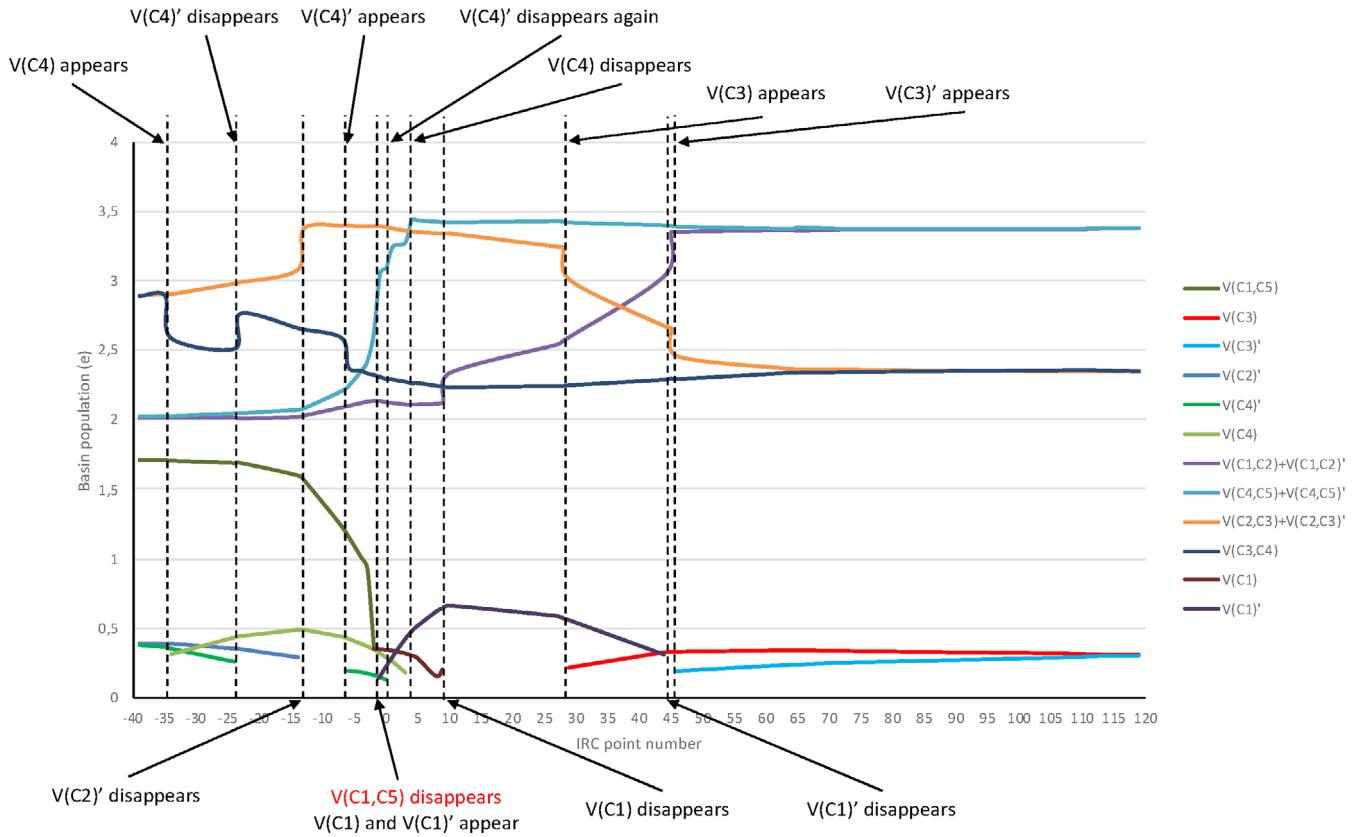


FIGURE 6 Basins population evolution along the IRC from **TS6**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The change reflecting the C1—C5 sigma bond disappearance is colored in red

As can be seen in Figure 6, the C1—C5 bond breaking takes place shortly before the TS of the process, although the population of the V(C1,C5) basin (olive green line) was decreasing from the beginning of the process, initially giving electrons mainly to the basins between C4 and C5 (light blue line), and finally to the transient monosynaptic basins on C1 (aubergine-colored and black lines).

On the other hand, the initial monosynaptic V(C4)' basin disappears quite early, and its population (green line) goes to the V(C3,C4) disynaptic basin (dark blue line), but it re-appears at the IRC reverse point 6th recovering population from the same V(C3,C4) basin, finally disappearing at the TS, but in this case its remaining population goes to the basins between C4 and C5. Also, it can be seen that a transient monosynaptic basin appears on C4, V(C4), taking its population (pale green line) from V(C3,C4) and giving its remaining population to the basins between C4 and C5, that attain a final population of ca 3.5 |e| very soon.

Other two transient monosynaptic basins can be found on C1. When they appear take their population mainly from V(C1,C5), as explained. When these monosynaptic basins disappear, their population is transferred to the basins between C1 and C2, that also attain a final population of ca 3.5 |e| (violet line).

The final monosynaptic basins on C3 (red and sky-blue lines) appear late along the IRC, taking population from the basins between C2 and C3. These V(C2,C3) and V(C2,C3)' basins were increasing their population (orange line) at the beginning of the process, taking the electrons mainly from V(C2)', but they lose population from the IRC point −15 on.

As happened before with the other **INT2** formation process, it is interesting to observe the final equilibration of the populations of the basins between C4 and C5 with the populations of the basins between C1 and C2 due to the final formation of two equivalent quasi-double bonds. Also, the populations of the basins between C3 and C4 and between C2 and C3 reach the same value (ca 2.4 |e|) at the end of the process because of the formation of two equivalent more-than-single bonds.

In Scheme 13, the curly arrows describing these electronic flows are shown.

3.7 | INT3 to INT4 through TS7

The IRC from **TS7** includes 14 points in the reverse direction to **INT3** and 18 points in the forward direction to **INT4**, see Scheme 14, so that the ELF function has been calculated for a total of 66 points. Along the IRC, 7 SSDs have been found (see Figure S11).

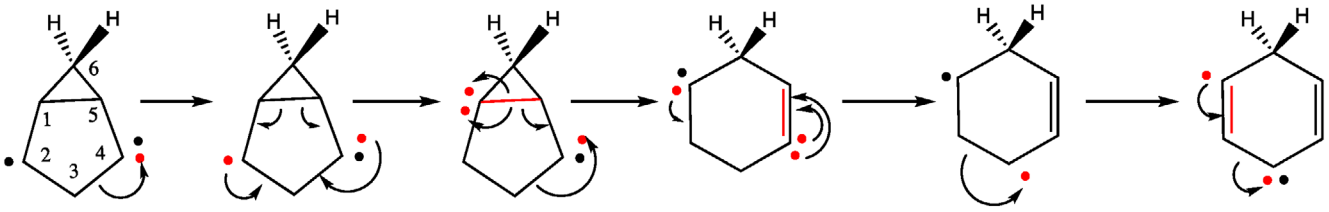
Among the 6 turning points, the more significant change refer to the disappearance of the V(C5,C6) disynaptic basin accounting for the C5—C6 bond breaking, that takes place at the TS, when SSD-V begins, with a C5—C6 distance of 2.022 Å. Other significant changes refer to the disappearance of the initial V(C2)' and V(C4)' monosynaptic basins taking place at the SSD-II and SSD-V, respectively, as well as to the appearance of the final V(C6) and V(C6)' monosynaptic basins that can be observed at the SSD-III and SSD-V, respectively.

As can be seen in Figure 7, the initial monosynaptic V(C2)' basin (orange line) disappears early, and its population goes to the disynaptic basins between C2 and C3, olive-green line. At the same time, a transient monosynaptic basin V(C4) appears, brown line, taking its population from V(C3,C4), red line. A little bit later, one of the final monosynaptic basins on C6 appear, violet line, taking its population from V(C1,C6), green line. From the beginning of the process, the population of V(C5,C6) is diminishing, blue line, populating the basins between C4 and C5, pale green line, as well as the basins V(C1,C6) and V(C1,C5), sky-blue line. Finally, at the TS the C5—C6 bond breaks, the second V(C6)' monosynaptic basin appears, light blue line, taking the remaining population of V(C5,C6), and both V(C4) and V(C4)' monosynaptic basins disappear, giving their populations to the basins between C4 and C5, that reach a total population of 3.3 |e|. In Scheme 15, the curly arrows describing these electronic flows are shown.

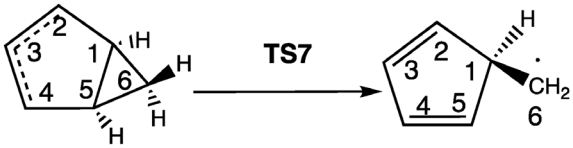
3.8 | INT4 to P2 through TS8

The IRC from **TS8** includes 43 points in the reverse direction to **INT4** and 25 points in the forward direction to **P2**, see Scheme 16, so that the ELF function has been calculated for a total of 138 points. Along the IRC, 6 SSDs have been found (see Figure S12).

Among the five turning points, the more significant changes refer to the disappearance of the V(C1,H) basin accounting for the C1—H bond breaking, the appearance of the V(H) basin for the leaving hydrogen, and also to the disappearance of the initial monosynaptic basins on C6. The first one takes place in the alpha part of the wavefunction when SSD-II is reached, while in the beta part it takes place when SSD-VI is reached, with a C1—H distance of 2.571 Å. The V(H) appearance takes place at SSD-II, with a C1—H distance of 1.706 Å, while the disappearances of V(C6) and V(C6)' take place at the SSD-II and SSD-III, respectively.



SCHEME 13 Main electron density flows for the **INT3** to **INT2** step



SCHEME 14 **INT3** to **INT4** step with atom numbering

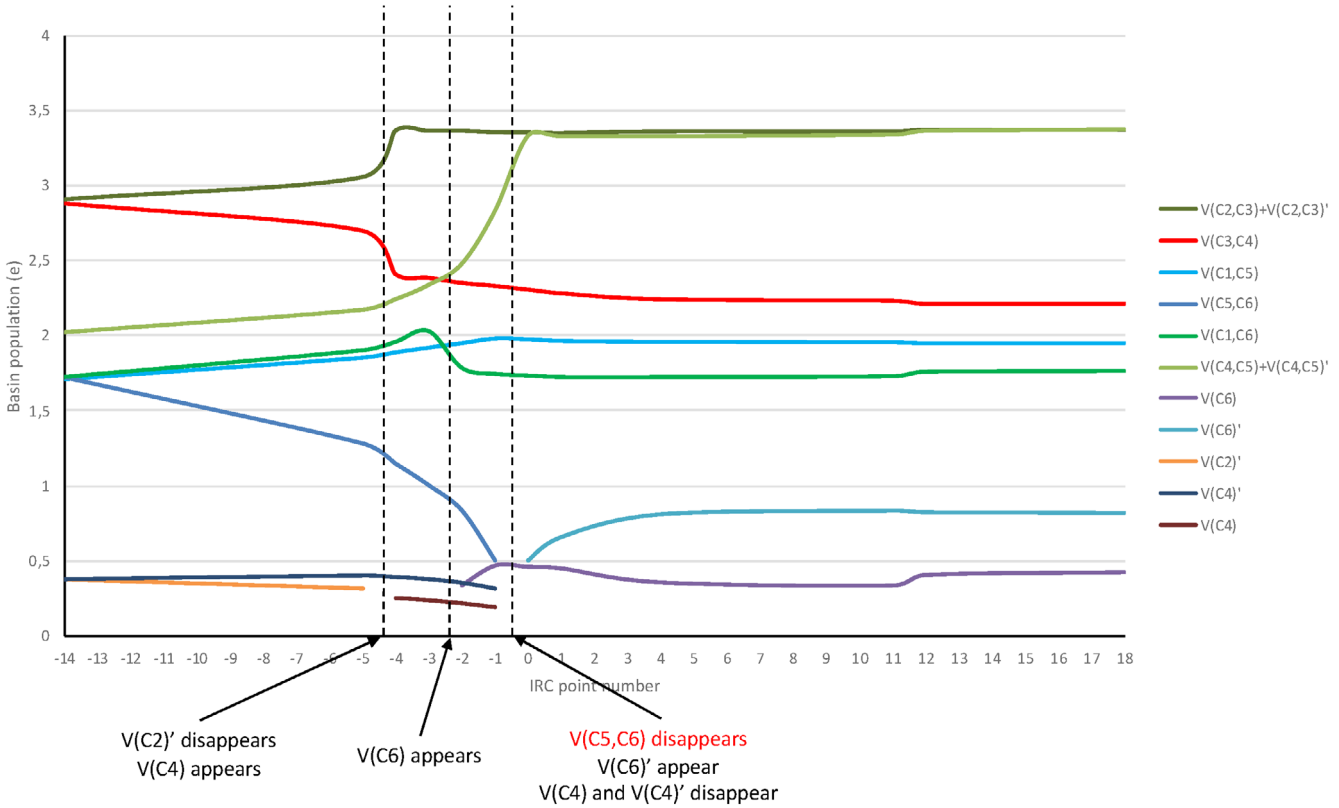
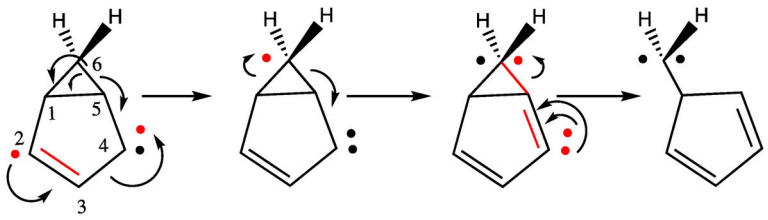


FIGURE 7 Basins population evolution along the IRC from **TS7**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The change reflecting the C5—C6 sigma bond disappearance is colored in red



SCHEME 15 Main electron density flows for the **INT3** to **INT4** step

SCHEME 16 **INT4** to **P2** step with atom numbering

As can be seen in Figure 8, the initial monosynaptic V(C6)' basin disappears at the same time that the V(H) basin appears. However, the population from V(C6)' goes to the disynaptic basins between C1 and C6, while V(H) takes its population from the former V(C1,H) basin. Shortly after that, the monosynaptic V(C6) basin also disappears, giving its population also to the basins between C1 and C6. Finally, the population of V(C1,H) diminishes, populating V(H) as well as the basins between C1 and C6, until the basin disappears accounting for the disappearance of the interactions between C1 and the leaving hydrogen. In Scheme 17 the curly arrows describing these electronic flows are shown.

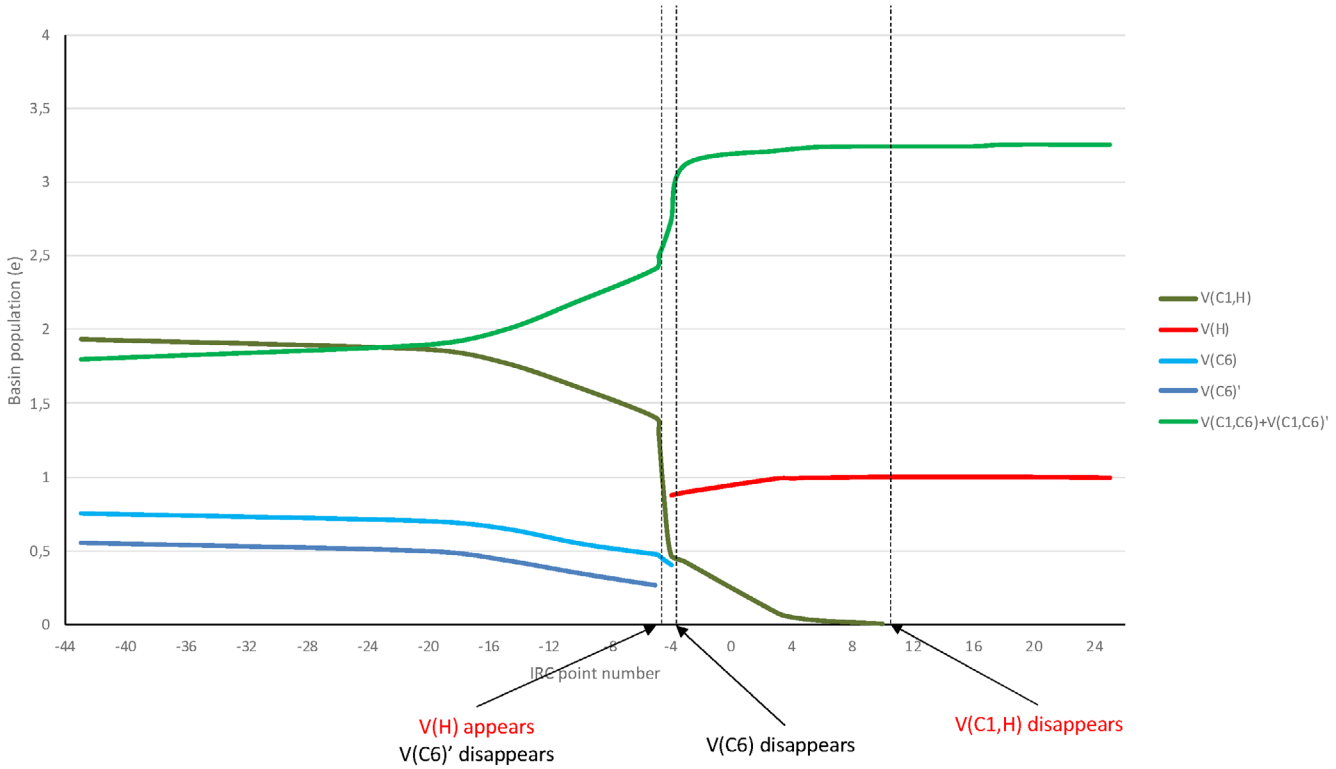
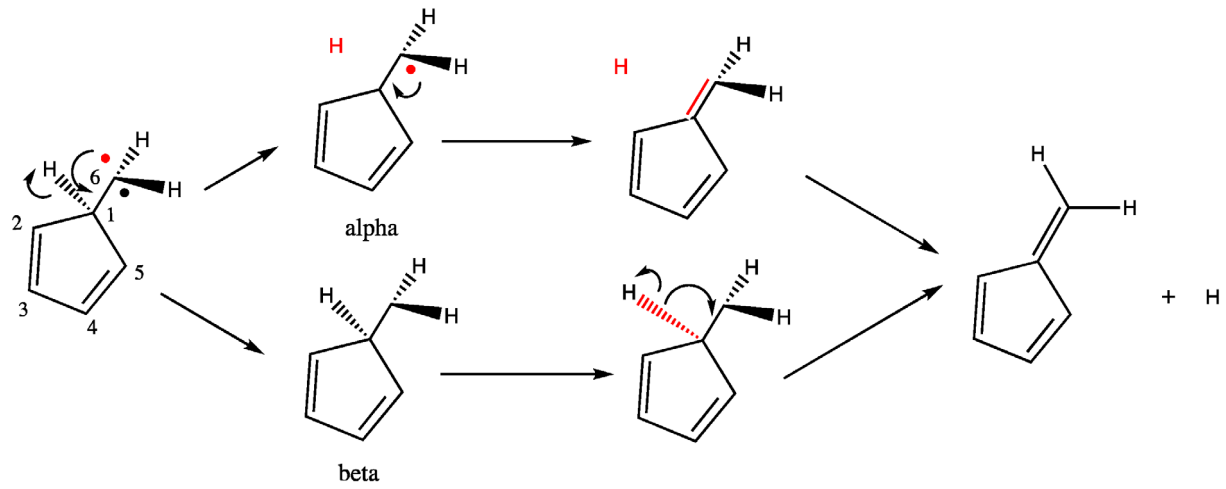


FIGURE 8 Basins population evolution along the IRC from **TS8**. Only selected turning points are indicated with dashed vertical lines, and the topological changes taking place at each one of these turning points are explained. The changes related with the C1—H breaking are colored in red



SCHEME 17 Main electron density flows for the **INT4** to **P2** step

4 | CONCLUSIONS

By using BET, that combines the analysis of the electron localization function and Thom's catastrophe theory, we report an exhaustive study of the cycloaddition reaction between the CH ($X^2\Pi$) radical and cyclopentadiene as explicative example of the BET use with open-shell systems. Our study is based on the calculation of the ELF function for a total of 1468 points along eight IRCs accounting for the different steps proposed to be opened in this complex process, that can render up to three different reaction products, namely benzene, fulvene and (*S*)-bicyclo[3.1.0]hexa-1,-3-diene. A complete picture on these pathways has been presented, and we have been able to characterize, at an electronic level, the processes responsible for the chemical transformations between the reactants and products along the different reaction intermediates, and to describe the

electronic flows in which the mechanism steps are based. Furthermore, we have also been able to massively apply the wavefunction splitting technique [7h] described in the Appendix, that is useful for the not very usual topological studies of systems with unpaired electrons, like the one herein studied, proving its value for this kind of analysis.

The present study allows us to conclude that in the steps from **INT1A** or **INT1B** to **INT2**, the breaking of the former sigma bond between C1 and C5 leads the electronic rearrangement. However, in the **INT3** to **INT2** process the breaking of the former C1—C5 sigma bond takes place after a previous electronic reorganization, because although in the three cases a bicyclic strained system relaxes to a six-membered cyclic system, the difference in the initial electronic distribution of **INT3** with respect **INT1A** or **INT1B** causes the electron flows to be very different. On the other hand, the benzene plus hydrogen formation from **INT2** has been described as an initial weakening of the C2—H bond with a quasi-hydride formation and a final retro-transfer of electrons from the quasi-hydride to the C2—H bond. In the way to the fulvene product, the **INT3** to **INT4** step is not leaded by the breaking of the sigma bond between C5 and C6, but by previous electronic redistribution, being the C5—C6 breaking one of the last topological changes sensed, while neither the fulvene formation from **INT4**, nor the interesting H transfer described in the **INT3** formation from **INT1B**, could be explained without the wavefunction splitting technique herein exemplified.

Present findings prove that BET is an appropriate tool to dissect, at atomic level, the molecular mechanism of chemical reactions also in open-shell systems and makes possible to visualize the bond breaking/forming evolution, that is, the electron density flows, in the framework of quantum mechanics. We are convinced that this theoretical methodology will foster novel studies of complex reactions, as we exemplify in this work.

ACKNOWLEDGMENTS

The authors also thank the Universitat Jaume I for computational resources.

AUTHOR CONTRIBUTIONS

Juan Andrés: Investigation; writing – original draft. **Vicent Sixte Safont:** Investigation; writing – original draft. **Mónica Oliva:** Investigation; writing – original draft. **Kacee L. Caster:** Investigation; writing – original draft. **Fabien Goulay:** Investigation; writing – original draft.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the annex as well as in the supplementary material of this article.

ORCID

Juan Andrés https://orcid.org/0000-0003-0232-3957
Vicent S. Safont https://orcid.org/0000-0003-2709-4230
Mónica Oliva https://orcid.org/0000-0001-6651-7852
Kacee L. Caster https://orcid.org/0000-0001-6926-9209
Fabien Goulay https://orcid.org/0000-0002-7807-1023

REFERENCES

[1] F. Meng, Y. Li, D. Wang, *J. Chem. Phys.* **2021**, *155*, 224111.
[2] (a) S. Álvarez, *Angew. Chem., Int. Ed.* **2012**, *51*, 590. (b) A. Nandi, S. Kozuch, *Chem. Eur. J.* **2020**, *26*, 759. (c) A. Lakshminarayanan, *Resonance* **2010**, *15*, 51. (d) R. V. Williams, A. A. Shaffer, *Can. J. Chem.* **2017**, *95*(3), 334. (e) G. Bhattacharyya, *J. Chem. Educ.* **2013**, *90*(10), 1282. (f) G. Bhattacharyya, G. M. Bodner, *J. Chem. Educ.* **2005**, *82*(9), 1402. (g) G. Bhattacharyya, G. M. Bodner, *J. Res. Sci. Teach.* **2014**, *51*(6), 694.
[3] (a) P. L. A. Popelier, Quantum Chemical Topology: on bonds and potentials. in *Structure and Bonding: Intermolecular Forces and Clusters I*, Vol. 115 (Ed: D. J. Wales), Springer, Heidelberg **2005**, p. 1. (b) P. L. A. Popelier, Quantum Chemical Topology. In *The Chemical Bond II. Structure and Bonding* book series vol 170 (Ed. M. P. Mingos), Springer, Heidelberg **2016**, p. 71
[4] X. Krokidis, S. Noury, B. Silvi, *J. Phys. Chem. A* **1997**, *101*(39), 7277.
[5] (a) A. D. Becke, K. E. Edgecombe, *J. Chem. Phys.* **1990**, *92*, 5397. (b) B. Silvi, A. Savin, *Nature* **1994**, *371*, 683.
[6] R. Thom, Structural Stability And Morphogenesis: an outline of a General Theory of Models, CRC Press, Taylor & Francis Group, Boca Raton, Florida **2018**.
[7] (a) P. González-Navarrete, J. Andrés, S. Berski, *J. Phys. Chem. Lett.* **2012**, *3*(17), 2500. (b) V. Polo, J. Andres, *J. Comput. Chem.* **2005**, *26*(14), 1427. (c) V. Polo, P. Gonzalez-Navarrete, B. Silvi, J. Andres, *Theor. Chem. Acc.* **2008**, *120*(4–6), 341. (d) M. Oliva, V. S. Safont, P. González-Navarrete, J. Andrés, *Theor. Chem. Acc.* **2017**, *136*(4), 51. (e) J. Andrés, L. Gracia, P. Gonzalez-Navarrete, V. S. Safont, *Comput. Theor. Chem.* **2015**, *1053*, 17. (f) J. Andres, S. Berski, J. Contreras-Garcia, P. Gonzalez-Navarrete, *J. Phys. Chem. A* **2014**, *118*(9), 1663. (g) J. Andrés, P. González-Navarrete, V. S. Safont, *Int. J. Quantum Chem.* **2014**, *114*, 1239. (h) V. S. Safont, P. Gonzalez-Navarrete, M. Oliva, J. Andres, *Phys. Chem. Chem. Phys.* **2015**, *17*(48), 32358.
[8] J. Andrés, S. Berski, B. Silvi, *Chem. Commun.* **2016**, *52*, 8183.
[9] (a) J. Andres, P. Gonzalez-Navarrete, V. S. Safont, B. Silvi, *Phys. Chem. Chem. Phys.* **2017**, *19*, 29031. (b) P. González-Navarrete, J. Andrés, V. S. Safont, *Phys. Chem. Chem. Phys.* **2018**, *20*, 535.

[10] (a) S. Berski, J. Andres, B. Silvi, L. R. Domingo, *J. Phys. Chem. A* **2003**, *107*(31), 6014. (b) V. Polo, J. Andres, S. Berski, L. R. Domingo, B. Silvi, *J. Phys. Chem. A* **2008**, *112*(31), 7128. (c) J. Andres, S. Berski, L. R. Domingo, V. Polo, B. Silvi, *Curr. Org. Chem.* **2011**, *15*(20), 3566. (d) S. Berski, J. Andrés, B. Silvi, L. R. Domingo, *J. Phys. Chem. A* **2006**, *110*(51), 13939. (e) N. Gillet, R. Chaudret, J. Contreras-García, W. Yang, B. Silvi, J.-P. Piquemal, *J. Chem. Theor. Comput.* **2012**, *8*(11), 3993. (f) I. Viciano, P. González-Navarrete, J. Andrés, S. Martí, *J. Chem. Theor. Comput.* **2015**, *11*, 1470. (g) C. Narth, Z. Maroun, R. A. Boto, R. Chaudret, M.-L. Bonnet, J.-P. Piquemal, J. Contreras-Garcia, A Complete NCI Perspective: From New Bonds to Reactivity. in *Challenges and Advances in Computational Chemistry and Physics*, Springer, Heidelberg **2016**, p. 491.
[11] P. Swings, L. Rosenfeld, *Astrophys. J.* **1937**, *86*, 483.
[12] (a) N. Love, R. N. Parthasarathy, S. R. Gollahalli, *Int. J. Green Energy* **2011**, *8*(1), 113. (b) F. V. Tinaut, M. Reyes, B. Giménez, J. V. Pastor, *Energy Fuels* **2011**, *25*(1), 119.
[13] A. J. Trevitt, F. Goulay, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5867.
[14] K. L. Caster, Z. N. Donellan, T. M. Selby, F. Goulay, *J. Phys. Chem. A* **2019**, *123*, 5692.
[15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr.., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09*, *Revision B.01*, Gaussian Inc., Wallingford, CT **2010**.
[16] S. Noury, X. Krokidis, F. Fuster, B. Silvi, *Comput. Chem.* **1999**, *23*(6), 597.
[17] A. I. Adjieufack, I. M. Ndassa, I. Patouossa, J. K. Mbadcam, V. S. Safont, M. Oliva, J. Andrés, *Phys. Chem. Chem. Phys.* **2017**, *19*, 18288.
[18] K. L. Caster, T. M. Selby, D. L. Osborn, S. D. Le Picard, F. Goulay, *J. Phys. Chem. A* **2021**, *125*, 6927.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: J. Andrés, V. S. Safont, M. Oliva, K. L. Caster, F. Goulay, *Int. J. Quantum Chem.* **2022**, *122*(11), e26892. https://doi.org/10.1002/qua.26892

APPENDIX A

BET for open-shell systems: An algorithm to split the wavefunctions

In our work, all paths take place, as explained, in the doublet electronic state, so that an unpaired electron is always present. Therefore, the wave functions were split into the alpha and beta parts. The whole procedure we use to apply BET to this kind of systems is as follows:

1. As explained in the Computational Methods section, starting from each Transition State (TS) geometry, we perform both the forward and reverse IRC calculations with the usual gaussian09 options (for instance, by using the following command line: #p UB3LYP/6-311g(2d,d,p) irc=(calcfc,noeigentest,maxcyc=120,maxpoints=200,forward). In this way, we obtain the "IRC-output-file".

2. Then we extract the geometry of each IRC point from this output, by using the command: awk -j "script" <"IRC-output-file"> "input-wfn-file", where "script" is the following executable file:

```
BEGIN{j=0
}
{
 if ($1=="CURRENT" && $2=="STRUCTURE") {
  getline
  getline
  getline
  getline
  getline
  getline
  i=0
  while (NF!=1) {
    i++
    at[i]=$2
    if (NF==5) {
     x[i]=$3
     y[i]=$4
     z[i]=$5
     }
    else {
     x[i]=$4
     y[i]=$5
     z[i]=$6
     }
    getline
    }
    j++
    print "%mem=2000Mb"
    print "%chk=for-"j".chk"
    print " "
    print "#P B3LYP/6-311g(2d,d,p) scf=(qc,maxcycle=500) guess=mix NoSymm stable=opt out=wfn"
    print " "
    print "for-"j
    print " "
    print "0 2"
    for (k=1;k<=i;k++) {
     print at[k],"  ", x[k], y[k], z[k]
     }
        print " "
        print "for-"j".wfn"
        print " "
        print "--Link1--"

      }
    }
```

The options wrote in red should be changed according to the particular system studied and the theoretical level used. Note that this executable file is intended to be used with a forward IRC, this is the reason for the "for" in green.

3. The resulting "input-wfn-file" is a gaussian09 input file used for calculating the wavefunction at each IRC point. The last "--Link1--" appearing in it must be deleted, and then this Gaussian 09 calculation can be done, thus generating a wavefunction file for each IRC point.

4. Then we create a file named "list" that includes the names of the wfn files (by typing "ls *.wfn | sort -V > list"), and we arrange this file eliminating the .wfn extensions (by executing: %s/\.wfn/\ /g inside "list").

5. Just to maintain a coherent nomenclature, we also arrange each wfn file by putting its own name in the first line, by executing the following script:

```
cat list | while read name
do
echo "sed 's/ $name/$name.wfn/' $name.wfn > $name.tmp " >> scriptil.sh
done
echo EXECUTE SCRIPTIL
. ./scriptil.sh
sleep 5
echo NOW MOVE

cat list | while read name
do
mv $name.tmp $name.wfn
done
```

Until now, the procedure is the same whatever the system being open or closed shell.

6. Now, to split the wavefunctions into the alpha and beta part we prepare a script that is different for each case. Let us explain this with an example:

We want to split all the wavefunction files obtained from an IRC forward calculation. An extract of the first one of these files, with the line numbers indicated in the left column of the table, is as follows:

```
1     for-1.wfn
2     GAUSSIAN          43 MOL ORBITALS    248 PRIMITIVES      13 NUCLEI
3      C    1   (CENTRE  1)  -1.29409201  2.07500244  0.09384191  CHARGE =  6.0
4      H    2   (CENTRE  2)  -1.78707621  4.01522017  0.50127442  CHARGE =  1.0
5      C    3   (CENTRE  3)  -2.79994674  0.00221476  0.37494245  CHARGE =  6.0
6      H    4   (CENTRE  4)  -4.73087268  0.02639003  1.05219573  CHARGE =  1.0
7      C    5   (CENTRE  5)  -1.42944554 -2.24360381 -0.27820170  CHARGE =  6.0
8      H    6   (CENTRE  6)  -2.14736572 -4.15301711 -0.16344241  CHARGE =  1.0
9      C    7   (CENTRE  7)   1.28028001  1.24783152 -0.73394128  CHARGE =  6.0
10     H    8   (CENTRE  8)   2.19675183  2.32744718 -2.22977290  CHARGE =  1.0
11     C    9   (CENTRE  9)   1.01638731 -1.55564712 -1.01194079  CHARGE =  6.0
12     H   10   (CENTRE 10)   2.41432167 -2.75605407 -1.88439522  CHARGE =  1.0
13     C   11   (CENTRE 11)   2.96618217  0.48192174  1.41442789  CHARGE =  6.0
14     H   12   (CENTRE 12)   4.96554265  0.18201842  1.11153313  CHARGE =  1.0
15     H   13   (CENTRE 13)   2.19214090  0.26592037  3.29010768  CHARGE =  1.0
16    CENTRE ASSIGNMENTS   1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
17    CENTRE ASSIGNMENTS   1  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2
18    CENTRE ASSIGNMENTS   3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
19    CENTRE ASSIGNMENTS   3  3  3  3  3  3  3  3  3  3  3  3  4  4  4  4  4  4  4  4
20    CENTRE ASSIGNMENTS   5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5  5
21    CENTRE ASSIGNMENTS   5  5  5  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6
22    CENTRE ASSIGNMENTS   7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
23    CENTRE ASSIGNMENTS   7  7  7  7  7  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8
24    CENTRE ASSIGNMENTS   9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9  9
25    CENTRE ASSIGNMENTS   9  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10 10 10 10 10
26    CENTRE ASSIGNMENTS  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
27    CENTRE ASSIGNMENTS  11 11 11 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12
28    CENTRE ASSIGNMENTS  13 13 13 13 13 13 13 13
29    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
30    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
31    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
32    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
33    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
34    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
35    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
36    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
37    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
38    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
39    TYPE ASSIGNMENTS     1  1  1  1  1  1  1  1  1  2  2  2  3  3  3  4  4  4  1  2
40    TYPE ASSIGNMENTS     3  4  1  2  3  4  5  6  7  8  9 10  1  1  1  1  1  2  3  4
41    TYPE ASSIGNMENTS     1  1  1  1  1  1  2  3  4
42    EXPONENTS  0.4563240D+04 0.6820240D+03 0.1549730D+03 0.4445530D+02 0.1302900D+02
...   ...
91    EXPONENTS  0.7500000D+00 0.7500000D+00 0.7500000D+00
92    MO    1    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.188153
...   ...
143   MO    2    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.186220
...   ...
194   MO    3    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.184113
...   ...
245   MO    4    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.168711
...   ...
296   MO    5    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.167931
...   ...
347   MO    6    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.167271
...   ...
398   MO    7    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.879591
...   ...
449   MO    8    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.756194
...   ...
500   MO    9    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.705745
...   ...
551   MO   10    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.627496
...   ...
602   MO   11    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.550886
...   ...
653   MO   12    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.540709
...   ...
704   MO   13    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.522318
...   ...
755   MO   14    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.453393
...   ...
806   MO   15    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.418385
...   ...
857   MO   16    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.396412
...   ...
908   MO   17    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.381565
...   ...
959   MO   18    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.363410
...   ...
1010  MO   19    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.357112
...   ...
1061  MO   20    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.315107
...   ...
1112  MO   21    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.250118
...   ...
1163  MO   22    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.164874
...   ...
1214  MO  151    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.188500
...   ...
1265  MO  152    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.186409
...   ...
1316  MO  153    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.176847
...   ...
1367  MO  154    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.169726
...   ...
1418  MO  155    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.162888
...   ...
1469  MO  156    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =  -10.162372
...   ...
1520  MO  157    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.874353
...   ...
1571  MO  158    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.747491
...   ...
1622  MO  159    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.697326
...   ...
1673  MO  160    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.613704
...   ...
1724  MO  161    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.546744
...   ...
1775  MO  162    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.535972
...   ...
1826  MO  163    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.518724
...   ...
1877  MO  164    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.448179
...   ...
1928  MO  165    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.413751
...   ...
1979  MO  166    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.394287
...   ...
2030  MO  167    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.378196
...   ...
2081  MO  168    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.356163
...   ...
2132  MO  169    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.353445
...   ...
2183  MO  170    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.295236
...   ...
2234  MO  171    MO 0.0     OCC NO =    1.0000000  ORB. ENERGY =   -0.233287
...   ...
2285  END DATA
2286   THE  HF ENERGY =   -232.794929189616 THE VIRIAL(-V/T)=   2.00465816
```

The line numbers are shown because they are used it in the splitting script. As can be seen, in this case there are 22 Molecular Orbitals in the alpha part, numbered from 1 to 22, and 21 Molecular Orbitals in the beta part, numbered from 151 to 171.

The script used to split all the wavefunctions obtained from the IRC forward calculation, listed in the file named "list" as explained before, is:

```
rm script.sh
cat list | while read n
do
#beta part
echo "sed '92,1213d' $n.wfn>tmp0.wfn" >> script.sh
echo "sed 's/MO  151/MO    1/' tmp0.wfn > tmp1.wfn" >> script.sh
echo "sed 's/MO  152/MO    2/' tmp1.wfn > tmp2.wfn" >> script.sh
echo "sed 's/MO  153/MO    3/' tmp2.wfn > tmp3.wfn" >> script.sh
echo "sed 's/MO  154/MO    4/' tmp3.wfn > tmp4.wfn" >> script.sh
echo "sed 's/MO  155/MO    5/' tmp4.wfn > tmp5.wfn" >> script.sh
echo "sed 's/MO  156/MO    6/' tmp5.wfn > tmp6.wfn" >> script.sh
echo "sed 's/MO  157/MO    7/' tmp6.wfn > tmp7.wfn" >> script.sh
echo "sed 's/MO  158/MO    8/' tmp7.wfn > tmp8.wfn" >> script.sh
echo "sed 's/MO  159/MO    9/' tmp8.wfn > tmp9.wfn" >> script.sh
echo "sed 's/MO  160/MO   10/' tmp9.wfn > tmp10.wfn" >> script.sh
echo "sed 's/MO  161/MO   11/' tmp10.wfn > tmp11.wfn" >> script.sh
echo "sed 's/MO  162/MO   12/' tmp11.wfn > tmp12.wfn" >> script.sh
echo "sed 's/MO  163/MO   13/' tmp12.wfn > tmp13.wfn" >> script.sh
echo "sed 's/MO  164/MO   14/' tmp13.wfn > tmp14.wfn" >> script.sh
echo "sed 's/MO  165/MO   15/' tmp14.wfn > tmp15.wfn" >> script.sh
echo "sed 's/MO  166/MO   16/' tmp15.wfn > tmp16.wfn" >> script.sh
echo "sed 's/MO  167/MO   17/' tmp16.wfn > tmp17.wfn" >> script.sh
echo "sed 's/MO  168/MO   18/' tmp17.wfn > tmp18.wfn" >> script.sh
echo "sed 's/MO  169/MO   19/' tmp18.wfn > tmp19.wfn" >> script.sh
echo "sed 's/MO  170/MO   20/' tmp19.wfn > tmp20.wfn" >> script.sh
echo "sed 's/MO  171/MO   21/' tmp20.wfn > tmp21.wfn" >> script.sh

echo "sed 's/$n.wfn/beta$n.wfn/' tmp21.wfn>tmp22.wfn" >> script.sh
echo "sed 's/GAUSSIAN          43/GAUSSIAN          22/' tmp22.wfn>beta$n.wfn" >> script.sh
#alpha part
echo "sed '1214,2284d' $n.wfn>tmp23.wfn" >> script.sh
echo "sed 's/$n.wfn/alpha$n.wfn/' tmp23.wfn>tmp24.wfn" >> script.sh
echo "sed 's/GAUSSIAN          43/GAUSSIAN          21/' tmp24.wfn>alpha$n.wfn" >> script.sh

done
echo EXECUTE SCRIPT
. ./script.sh
sleep 5
echo NOW REMOVE
rm tmp*.wfn
```

In this way, two wavefunction files for the first IRC point, alphafor-1.wfn and betafor-1.wfn, are obtained due to the splitting of the for-1.wfn wavefunction. The same happens with the rest of files for this IRC forward calculation.

7. Once this is done, for each wavefunction the input files needed for the ELF calculations should be prepared and launched. We use the TopMod package, as explained in the Computational Methods section, and therefore, we follow the Bernard Silvi's manual instructions for preparing and launching the ELF calculations. The manual can be found here: https://www.lct.jussieu.fr/pagesperso/fuster/TOPMOD/topmod-manual.pdf.

8. This provides the files containing the ELF data for each alpha and beta part of every IRC point. The number and nature of basins and their populations are obtained for each case, so that the BET analysis can be conducted.