PCCP

Unveiling the Reaction Pathways of Hydrocarbon via Experiment, Computations and Data Science†

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-09-2022-004499.R1
Article Type:	Paper
Date Submitted by the Author:	14-Nov-2022
Complete List of Authors:	Takahashi, Lauren; Hokkaido University, Yoshida, Shigehiro; Honda R & D Co Ltd Fujima, Jun; Hokkaido University, Department of Chemistry Oikawa, Hiroshi; Honda R & D Co Ltd Takahashi, Keisuke; Hokkaido University, Department of Chemistry; National Institute for Materials Science, Center for Materials research by Information

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Unveiling the Reaction Pathways of Hydrocarbon via Experiment, Computations and Data Science†

Lauren Takahashi[a], Shigehito Yoshida[b], Jun Fujima[a], Hiroshi Oikawa*[b], Keisuke Takahashi*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Reaction networks of hydrocarbon are explored using first principles calculations, data science, and experiment. Transforming hydrocarbon data into networks reveals the prevalence of the formation and reaction of various molecules. Graph theory is implemented to extract the knowledge from reaction network. In particular, centralities analysis reveals that H+, C=CC, CH_3+, C=C, and [CH_2+]C have high degrees and are thus very likely to form or react with other molecules. Additionally, H+, CH_3+, C_2H_5+, C_8H_{15}+, C_8H_{17}+, and C_6H_{11}+ are found to have high control throughout the network and lead towards series of additional reactions. Constructed network is also validated in experiment while the shortest path analysis is implemented for further comparison between experiment and network. Thus, combining network analysis with first principles calculations uncovers key points in the development of various hydrocarbons that can be used to improve catalyst design and targetted synthesis of desired hydrocarbons.

1 Introduction

Hydrocarbon acts as a key combustible fuel source as well as a precursor for synthesizing polymeric materials[1–4]. For instance, aromatic compounds such as C_4 and C_8 are the main components of natural gas and gasoline, respectively while ethylene(C_2H_4) is a precursor for synthesizing polyethylene[5–7]. In order to control the combustion as well as the synthesis of particular hydrocarbon, catalysts as well as cracking are designed to eliminate pollutants or increase the yield of particular hydrocarbon[8–10]. However, the details behind the mechanism of how hydrocarbons transform from C_1 to C_8 or from C_8 to C_1 remains a mystery due to the complexity of the reaction[11]. Complexity in the hydrocarbon reaction lies within the existence of isomers among the hydrocarbon as well as complication of how such isomers interact each other[12]. Furthermore, it is uncertain what the roles of intermediates and isomers have within a hydrocarbon reaction. If the details of the relations of hydrocarbon are revealed, the whole picture of a hydrocarbon reaction network becomes visible and can thereby be used to help design catalysts that improve yield or for synthesizing target hydrocarbon.

Understanding the detailed hydrocarbon reaction network is a challenging matter due to the complex chemical and physical matter that arise due to the vast number of elementary reactions involved among the hydrocarbon and in reaction networks in general[13–16]. Although the reactants and products are observed in experiment, the lifetime of intermediate molecules are too short to capture in spectroscopies[17–19]. Given such circumstances, first principle calculations make it possible to unveil the atomic scale reactions where various automated transition state search algorithms have been proposed[20–22]. The aim of this work is to understand the complex reaction mechanism of C_1H_x-C_8H_x reactions, which can be helpful towards understanding pyrolysis, combustion, polymerization that involves C_{1-8} compounds, and hydrocarbon formation and activity under extraterrestrial scenarios. Here, various elementary reactions within hydrocarbon from C_1 to C_8 are explored by combining the first principle calculations and automated transition state search algorithm in order to reveal the whole pictures of hydrocarbon reaction network. The results are then analyzed and validated by data science and experiment, respectively.

2 Method

2282 elementary reactions within hydrocarbon from C_1 to C_8 are calculated. Data consisting of C_1H_x to C_8H_x, related molecules, and their reactions are calculated and collected using Kinbot[23].

[a] Department of Chemistry, Hokkaido University, North 10, West 8, Sapporo 060-0810, Japan
* email:keisuke.takahashi@sci.hokudai.ac.jp
[b] Innovative Research Excellence, Power unit Energy, Honda R&D Co., Ltd., 3-15-1 Senzui, Asaka, Saitama, 351-0024, Japan
* email:hiroshi_oikawa@jp.honda

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

Kinbot is performed using the b3lyp function and 6-31+g and is used to automatically search and calculate the barrier. In particular, Kinbot with Gaussian 16 is implemented in order to calculate the transition state energy from reactants to products[23]. Molecules are written using SMILES notation and are then converted to molecular formula using Python libraries OpenBabel and RDKit for further investigation[24,25]. This data is then preprocessed and used to construct a reaction network for further analysis using Gephi[26]. In the network, reactants and products are set to nodes while transition state energy is set as edges. Shortest paths are calculated using NetworkX and are used when comparing the findings within the network against experimental observations[27].

Experiments are carried out for comparison purposes where reactions of C_2H_4, 1-C_4H_8, cis-2-C_4H_8, trans-2-C_4H_8, and iso-C_4H_8 are performed with SiO_2. SiO_2 is selected as the support as it is considered to be stable and unreactive due to the strong bond between Si and O. Given this, one can consider that the effect of SiO_2 within experiments should be minimum and is thus used to control gas flow. SiO_2-CARiACT G-6 [Particle size : 10μm]- is purchased from FUJI SILYSIA CHEMICAL LTD where the weight of SiO_2 is 0.3g. Gas feed rate of C_2 and C_4 is 10mL/min with the pressure of 0.1 MPa. Catalysts are placed in SUS316 reactor with a reactor diameter of 3/8 inch. C_2H_4, 1-C_4H_8, cis-2-C_4H_8, trans-2-C_4H_8, and iso-C_4H_8 conversion and selectivities of C1-C_{12} are measured at temperatures of 200°C, 300°C, 400°C, 500°C, and 600°C.

3 Results and Discussion

Data is preprocessed and transformed into networks in order to better understand the development and cracking of various hydrocarbons. While Kinbot searches the decomposition of single molecules, it is possible to connect these molecules via network in order to better visualize the series of reactions that may occur during hydrocarbon cracking. For instance, C_8H_{17} can be decomposed to molecules such as C_5H_{10}, C_3H_6, and C_2H_4. It is also possible to consider how molecules may form by considering the inverse TS. These molecules can then be connected together via network where molecules are represented as nodes in the network and their connections are represented as edges.

For visualization purposes, the following are extracted and used when preparing data for network visualization: Reaction, Product 1, Product 2, TS, and TS Inverse. Note that TS represents the transition state energy required in order for a reaction to occur while TS Inverse represents the energy mountain required for the inverse reaction to occur. Reaction, Product 1, and Product 2 are used to create network nodes while TS and TS Inverse are used to determine edge weights.

Figure 1 illustrates the resulting network of C_8-related reactions according to their SMILES notation, which makes it possible to consider isomers when studying the reaction network. The network illustrated in Figure 1 is composed of 397 nodes and 3,322 edges. Note that the molecules considered may not cover all possibilities that may occur throughout the reaction due to the high complexity of hydrocarbons. Nodes represent molecules and are colored grey. Nodes that are colored yellow, meanwhile, denote the top 10 nodes with the largest in-degrees or out-degrees where degree represents the number of edge connected to the node. Label colors distinguish nodes that are found to be nodes with different highest centralities. In particular, blue labels denote nodes with top 10 betweenness centrality and magenta labels denote nodes with top 10 closeness centrality. Green labels denote nodes with both top 10 betweenness centrality and top 10 closeness centrality. Edge colors reflect weight, where red-colored edges have smaller weights and are less likely to occur in comparison to blue-colored edges, which have higher weights and are thus more likely to occur. Weight reflects TS, which reflects the energetic mountains that must be overcome before a reaction occurs. In order to accurately reflect that smaller TS values have higher weights in Gephi, TS values had to be edited and inversed in order for the ForceAtlas2 algorithm within Gephi to accurately reflect this relationship in the network visualization. Original TS values and the edited values used for the visualization purposes are both included within Supporting Information.

At first glance, one can see that groups of nodes cluster together with several nodes appearing to act as pivot points between said node clusters. Figure 4 illustrates the central portion of the network, where it can be seen that nodes H+, C=CC, CH_3+, C=C, and [CH_2+]C share edges with a very large number of nodes. In another word, one can considered that those molecules in Figure 4 are important intermediates in the hydrocarbon reactions. Table S1, found within Supporting Information, lists the top 10 molecules with the highest degrees. Analysis of in-degrees and out degrees confirm that these molecules and others have high degrees, suggesting that these molecules are very likely to result from various reactions or are very likely to react and lead towards the formation of other molecules. Edge color helps visualize which groups of molecules have large TS, which suggests that the reaction requires a lot of energy in order to occur, or small TS values, which suggests that the reactions requires less energy in order to occur. From Figures 1 and 4, one can see that reactions involving H+ are very likely to occur. Edge color also helps visualize which isomers are more likely to form, which is denoted by the dark blue colors of the edges. Thus, by visualizing the various hydrocarbon data into a network, it becomes easier to see which molecules are more likely to occur or are likely to have a key role.

Betweenness centrality is analyzed in order to better determine the number of connections nodes share within the network. In particular, betweenness centrality is of interest as nodes with high betweenness centrality are often understood to have high control over the network. This allows one to better understand which nodes have a stronger influence over the formation or location of various other nodes throughout the network. Specific betweenness centralities are listed in Table 1 and accompanying SMILES structures are illustrated in Figure 2.

From Table 1, one can see that H+, CH_3+, C_2H_5+, C_8H_{15}+, C_8H_{17}+, and C_6H_{11}+ experience the highest number of connections within the network. This suggests that these nodes have a high level of control within the network. A closer look at Figure 4 helps illustrates this. To start, in Figure 4 one can see that nodes with high betweenness (which are labeled in blue or green) can

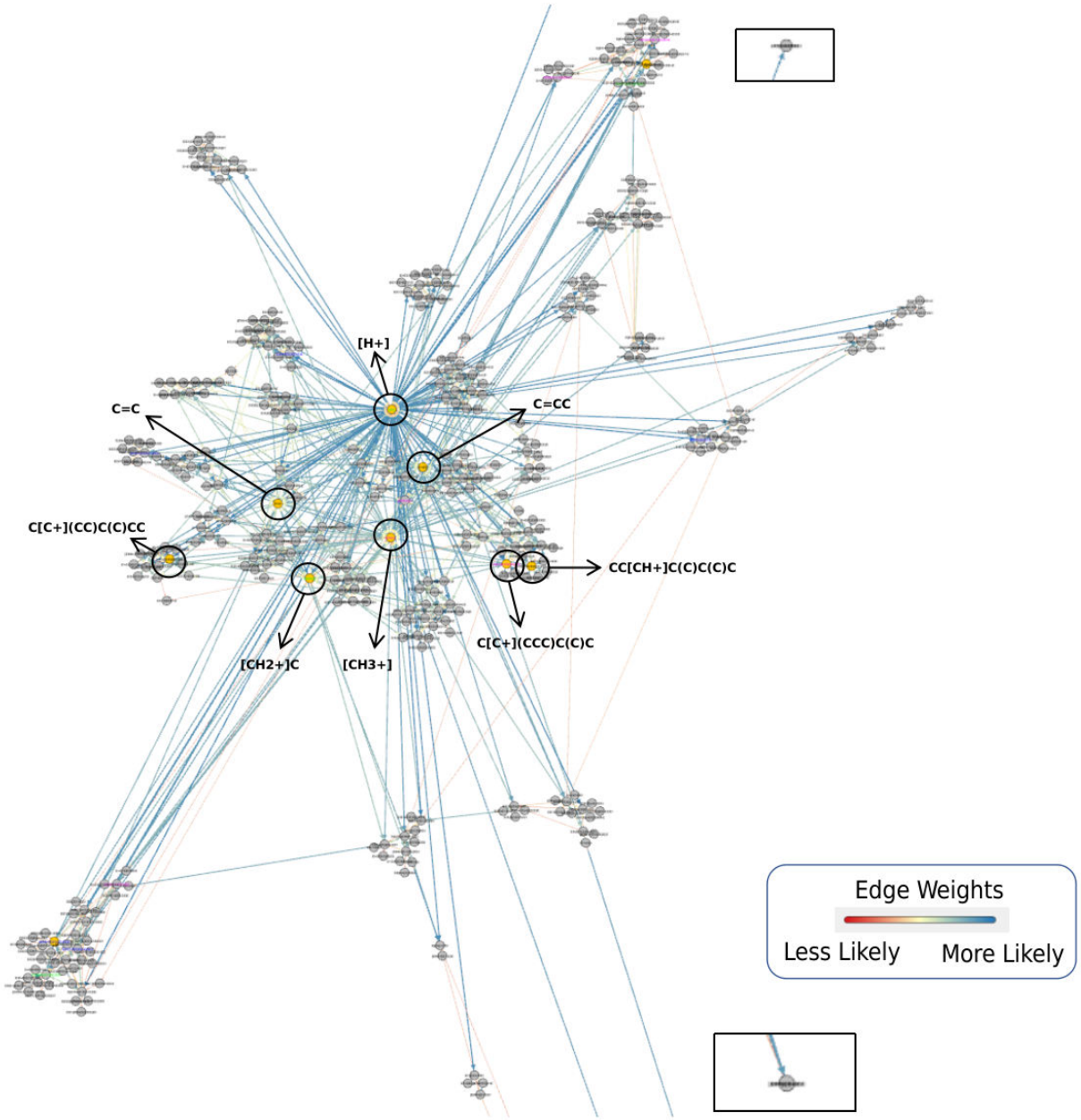

Fig. 1 Hydrocarbon reaction network. SMILES notation is used.

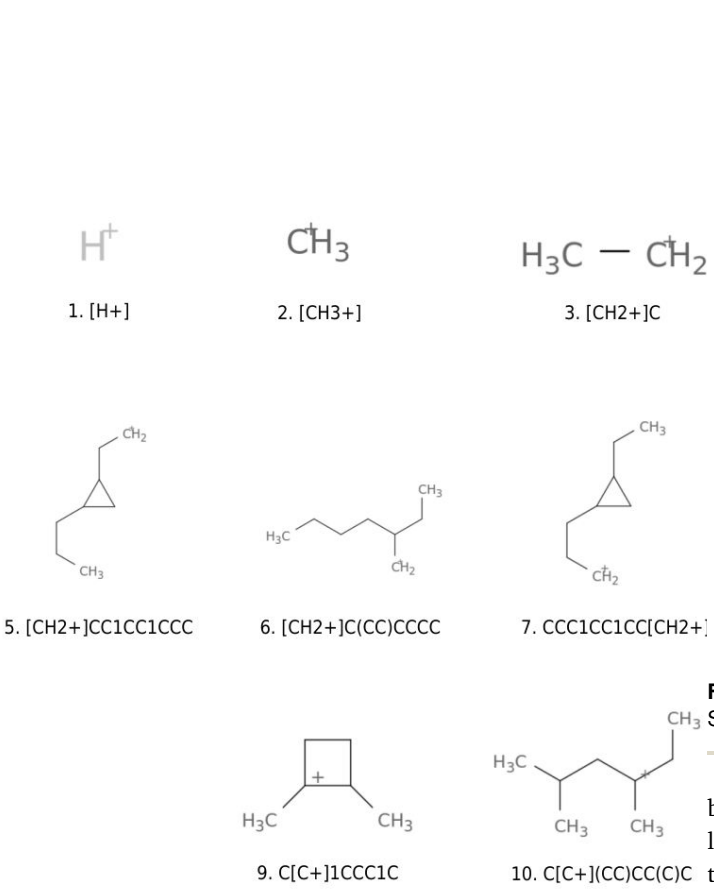

Fig. 2 The nodes with Top 10 Betweenness Centrality and accompanying SMILES notation.

A	B	C
[H+]	H+	91244.40
[CH3+]	CH_3+	37693.71
[CH2+]C	C_2H_5+	7887.10
CC[CH+]C1CC1CC	C_8H_{15}+	6329.07
[CH2+]CC1CC1CCC	C_8H_{15}+	5444.85
[CH2+]C(CC)CCCC	C_8H_{17}+	5361.35
CCC1CC1CC[CH2+]	C_8H_{15}+	5092.40
C[C+]1C(C)C1CCC	C_8H_{15}+	4877.90
C[C+]1CCC1C	C_6H_{11}+	4415.23
C[C+](CC)CC(C)C	C_8H_{17}+	4341.38

Table 1 Top 10 Betweenness Centralities. A: SMILES formula. B: Molecular Formula. C: Betweenness. These nodes correlate with the blue-labeled nodes found in Figures 1 and 4.

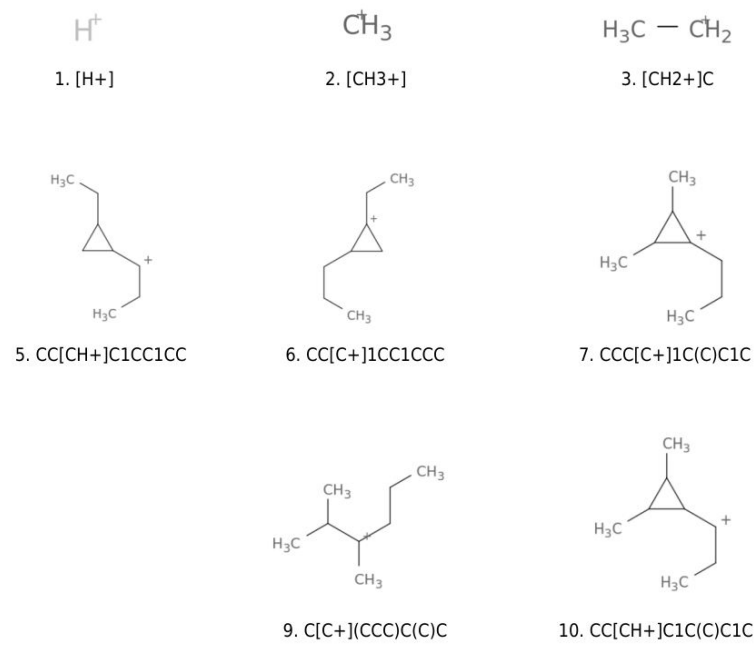

Fig. 3 The nodes with Top 10 Closeness Centrality and accompanying SMILES notation.

be found at the centers of various clusters of nodes. Given their locations within the network, one can come to the understanding that these molecules could be important intermediates for many reactions that may occur when breaking down octane. Upon converting the listed SMILES formulas to their molecular formulas, it becomes more clear that C_8H_{17}+ and C_8H_{15}+ in particular play central roles where these molecules either experience changes in their structure and become different isomers or lead to the formation of other molecules. Interestingly, by utilizing SMILES notation, it becomes possible to trace the paths that may lead to particular isomers of a given molecule. Thus, analyzing betweenness centrality allows one to determine which molecules play key roles and lead towards the formation of different sets of molecules.

Closeness centrality is then analyzed in order to better understand which nodes experience the highest level of traffic within the network. Understanding closeness centrality can thereby help unveil which nodes are the most central, or closest, to all other nodes within the network. This allows one to better understand which nodes may be more likely to be encountered when calculating the shortest paths between two nodes. Specific closeness centralities for the top 10 cases are listed in Table 2 and accompanying SMILES structures are illustrated in Figure 3.

From Table 2, one can see that H+, CH_3+, and C_2H_5+ are within the top 3 molecules with highest closeness centrality, as also seen with the case of betweenness centrality. These molecules are then followed by C_3H_7+, several instances of C_8H_{15}+, and C_8H_{17}+. From this list, one can begin to understand that these molecules play a very important role regarding hydrocarbon-related reactions. To start, molecules H+, CH_3+, C_2H_5+, C_3H_7+, and C_8H_{17}+ are located at the center of the network. Their network locations, in combination with their high closeness centralities, suggest that these molecules are involved

A	B	C
[H+]	H+	0.54
[CH3+]	CH_3+	0.47
[CH2+]C	C_2H_5+	0.41
[CH2+]CC	C_3H_7+	0.39
CC[CH+]C1CC1CC	$C_8H_{15}+$	0.39
CC[C+]1CC1CCC	$C_8H_{15}+$	0.39
CCC[C+]1C(C)C1C	$C_8H_{15}+$	0.38
C[C+]1C(C)C1CCC	$C_8H_{15}+$	0.38
C[C+](CCC)C(C)C	$C_8H_{17}+$	0.38
CC[CH+]C1C(C)C1C	$C_8H_{15}+$	0.38

Table 2 Top 10 Closeness Centralities. A: SMILES formula. B: Molecular Formula. C: Closeness Centrality. These nodes correlate with the magenta-labelled nodes found in Figures 1 and 4.

in a particulary large number of reactions. H+, CH_3, C_3H_7, and C_2H_5, in particular, are not only central to the network and have very high numbers of edges that connect them to other molecules, but many of these edges are found to have higher edge weights. This, therefore, suggests that these molecules are actively involved in the formation of many different types of hydrocarbons and have a high probability of occuring. $C_8H_{15}+$ is also interesting in regards to high closeness centrality, as five of the top 10 molecules with the highest closeness centralities are isomers of C_8H_{15}. This suggests that C_8H_{15} has an even higher likelihood of not only contributing to the formation of other hydrocarbon molecules, but also in forming as the result of other similar reactions. Thus, by using closeness centrality it becomes possible to pinpoint which molecules are most likely to be involved in the formation of various hydrocarbons.

When analysing betweenness centrality and closeness centrality, it become clear that several molecules have particularly important roles within the network. These molecules– H+, C_2H_5+, and $C_8H_{15}+$– not only have top betweenness centrality but also top closeness centrality. In fact, these molecules not only are located near the center of the network- suggesting that they are highly likely to connect to a majority of the molecules within the network– but are also mainly involved with edges with high edge weights. H+, for example, shares the highest number of connections with other molecules where most of them have high edge weights and direct to other molecules. This suggests that H+, therefore, is very reactive and is most likely involved in the formation of other molecules. C_2H_5+ also has a high number of edges that connect to it, but in this case is most likely to be the result of other reactions. Finally, the sheer number of $C_8H_{15}+$ isomers present throughout the network- which all are passed through before forming other molecules- suggests that $C_8H_{15}+$ is highly likely to form and consequently lead to the formation of other molecules. Additionally, by representing the molecules by their SMILES notations, it becomes possible to pinpoint particular isomers of $C_8H_{15}+$ and what may result from their reactions, thus providing a more detailed look into how hydrocarbons form and react.

An additional network is constructed where the molecules are represented by their molecular formulas instead of their SMILES notation in order to better understand which molecules form and lead to the formation of other molecules. This network is illustrated in Figure 5 and provides a clearer picture of what molecules play important roles. Here, nodes with blue labels are nodes with top 10 betweenness centrality, nodes with pink labels are nodes with top 10 closeness centrality, and red labels are nodes that have both top 10 betweenness centrality and top 10 closeness centrality. The top 6 nodes with the highest in and out degrees are colored in yellow.

Figure 5 provides a clearer picture of what molecules play important roles. To start, one can see that H+ is located at the center of the network. Given its location and the directionality of its edges, it is safe to conclude that H+ is not only the product of many reactions, but is also highly likely to lead into the formation of other molecules.

In addition to H+, four hydrocarbons appear to be key to the development of further hydrocarbons: $C_8H_{17}+$, $C_8H_{15}+$, $C_6H_{11}+$, and $C_6H_{13}+$. In Figure 5, these molecules are located in different areas where groups of other molecules cluster around them. Edge arrows, which designate the direction in which the target node resides, are found to largely point to these nodes, suggestions that these molecules are very likely to form. In particular, a closer look at edge color suggests that H+ is much more likely to participate in the formation of molecules such as $C_6H_{13}+$, C_2H_5+, $C_8H_{15}+$, and $C_8H_{17}+$. In addition, edges that connect nodes to themselves appear for $C_8H_{15}+$, $C_6H_{11}+$, C_4H_7+, $C_6H_{13}+$, C_3H_7+, $C_8H_{17}+$, and C_4H_9+. This suggests that these molecules are also likely to change structure and form different isomers of itself. Thus, by transforming SMILES notation into molecular formulas and representing this data as a network, it becomes easier to understand hydrocarbons that play key roles during cracking.

Transforming SMILES notation into molecular formulas also allows one to better understand how particular molecules may form. For instance, it becomes easier to determine whether C_8 would be likelier to break down into C_3+C_5 or into C_4+C_4. In the network illustrated in Figure 5, there are a total of 104 reactions that lead to C_3+C_5 while there are 56 reactions that lead to C_4+C_4. This suggests that $C_8 \rightarrow C_3+C_5$ is likelier to occur than $C_8 \rightarrow C_4+C_4$. Additional analysis shows that 634 reactions involve C_6 either decomposing or changing isomers, whereas 1772 reactions involve the isomer change or decomposition of C_8.

This approach can also be taken when considering whether C_6 decomposes to C_3+C_3 or C_2+C_4. In this case, there are 24 reactions where C_6 leads to C_3 while there are 44 reactions that lead to C_2+C_4. Interestingly, further investigation into reactions where C_3, C_2, and C_4 are the source molecules reveals their likelihood to change isomers or lead to the formation of other molecules. In the case of C_3, 89 reactions occur where C_3 undergoes an isomer change or absorbs hydrogen. Meanwhile, 244 reactions occur where C_2 and C_4 are the source molecules; however, not only are they likely to undergo isomer change, but they also lead to the formation of C_6 and C_8 molecules. This suggests that C_2 and C_4 may continue to react upon formation, thereby making it more difficult to detect.

Finally, by studying the network, it is also possible to begin to understand how a certain molecule may form. For instance, C_8

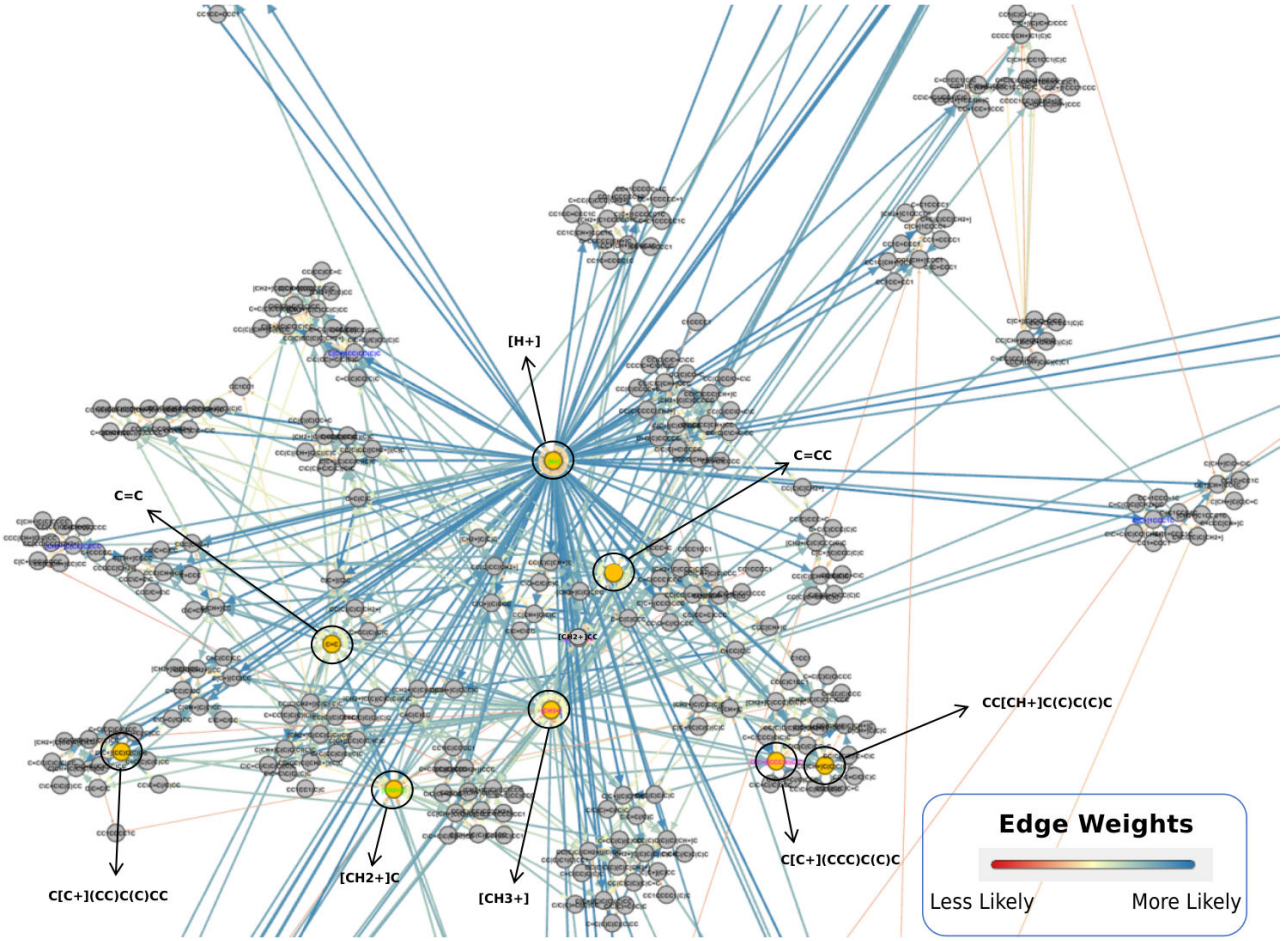

Fig. 4 The central section of the network illustrated in Figure 1.

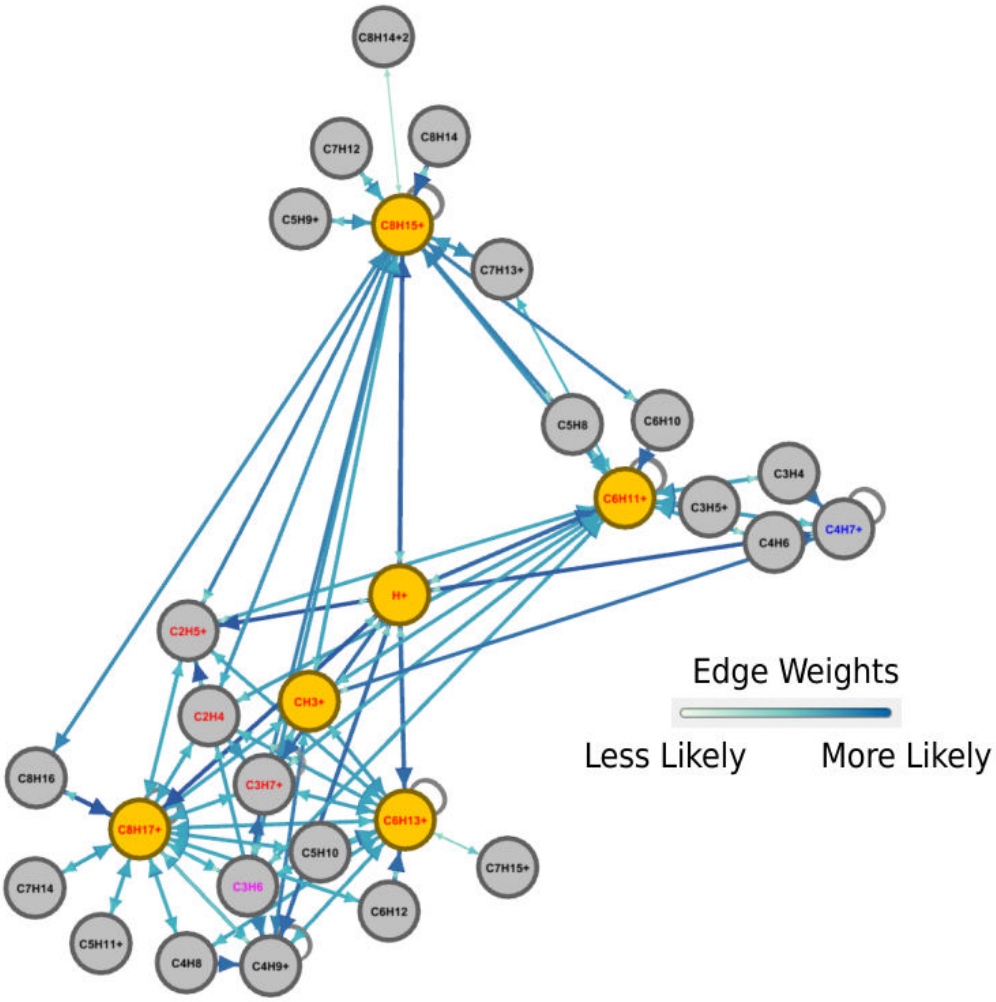

Fig. 5 The hydrocarbon network translated from SMILES notation to molecular formulas.

may be formed via C_6+C_2 or C_4+C_4. Network analysis reveals that 120 reactions related to C_6+C_2 result in the formation of C_8 while 56 reactions related to C_4+C_4 result in the formation of C_8. Given this, it is more likely that C_8 may result from C_6+C_2 reactions.

Representing hydrocarbon reaction data as networks lends insight into potential pathways that lead towards the formation of other notable molecules. For instance, the network illustrated in Figure 5 shows that C_2H_4 appears as a product of reactions involving molecules such as $C_8H_{15}+$, $C_8H_{17}+$, and $C_6H_{11}+$. Additionally, it becomes possible to understand that once C_2H_4 is formed, it is likely to form C_2H_5 due to the high weight of the connecting edge. Observing the positions of high edges throughout the network also demonstrate that H+ is very likely to react with other molecules. While edges leading from H+ towards other molecules have high weights, molecules such as C_8H_{16}, C_3H_6, and C_4H_6 also have high weight edges that connect to $C_8H_{17}+$, C_3H_7+, and C_4H_7+, respectively. This suggests that H+ is very reactive with other molecules, and most likely must be monitored when attempting to produce specific molecules.

In light of these observations, experiments are conducted in order to compare experimental observations with the information that can be extracted from the networks. Parallel coordinates are calculated for the resulting experimental results in order to better visually understand what molecules are most likely to be detected. The resulting plots are collected and reported in Supplementary Information as Figures S1, S2, S3, S4, and S5. Here, reaction of C_2H_4,1-C_4H_8, trans-2-C_4H_8, cis-2-C_4H_8, iso-C_4H_8 are performed as the increase of temperature. In particular, conversion of each molecules as well as the percentage of individual molecules detected from the total amount of molecules produced (100 %) after the reaction of those molecules are plotted. From these plots, it becomes easier to understand which molecules are detected at various temperatures.

Figure S1 plots the percentage of various hydrocarbons detected at various temperatures for reaction of C_2H_4 (C=C). From this, one can see that C=C remains the primary molecule detected until temperature reaches 600°C, where CC is primarily detected, followed by C and C=C. This suggests that C=C remains a stable structure until temperature reaches 600°C, where it then breaks down into other molecules. Meanwhile, Figure S2 plots the reaction of 1-C_4H_8 (C=CCC). Conversion ranges between 65 and 80 percent when temperature is within the range of 300-600°C. While C=CCC is the primary molecule detected at 200°C, it is reduced and C\C=C\C and C\C=C/C are largely detected as temperature increases. Thus, 1-C_4H_8 is found to lead towards the formation of other molecules as temperature increases.

Figure S3 plots the reaction of cis-2-C_4H_8 (C\C=C/C). Here, one can see that C\C=C/C is primarily detected when temperature is 200°C to 300°C. As temperature increases, detection of C\C=C/C is reduced as C=CC, C=CCC, and C\C=C\C are increasingly detected as temperature increases. This behavior reflects what is previously observed with 1-C_4H_8, where C\C=C/C is reduced as other molecules form with the increase in temperature. Figure S4 plots the reaction of trans-2-C_4H_8 (C\C=C\C). Conversion of trans-2-C_4H_8 ranges between 40 to 60 percent for temperatures 400°C to 600°C. Additionally, similar to previous observations, C\C=C\C is the primary detection while temperature 200°C and 300°C. At temperature 400°C, C\C=C\C detection is reduced and CCCC and C\C=C/C are also detected. As temperature continues to increase to 500°C and 600°C, C=CCC and C\C=C/C detection continues to increase while C\C=C\C detection decreases and CCCC is reduced. Meanwhile, Figure S5 plots the reaction of iso-C_4H_8 (C=C(C)C) is low across all temperatures where C=C(C)C is primarily detected. This sugggests that the structure of iso-C_4H_8 does not change with the increase of temperature to 600°C. Thus, from these graphs, one can understand that CC, C=C, C=CC, C=CCC, C\C=C\C, and C\C=C/C are primarily detected experimentally.

The prominence of these molecules, as seen via experiment, are also present within the constructed networks. Figures S6-S10 provide more-detailed views of the reaction networks illustrated in Figures 1 and 4. As can be seen in these figures, the aforementioned molecules are located in areas that are near the center of the network, suggesting that the nodes representing these molecules play important roles in the composition and decomposition of various hydrocarbons. To follow these results, additional calculations are carried out using NetworkX where the shortest paths between C_2H_4, 1-C_4H_8, cis-2-C_4H_8, trans-2-C_4H_8, and iso-C_4H_8 and select molecules for each instance are calculated. The resulting paths are listed in Tables 3, 4, 5, 6, and 7. Source nodes are set to C_2H_4, 1-C_4H_8, cis-2-C_4H_8, trans-2-C_4H_8, and iso-C_4H_8– represented by C=C, C=CCC, C\C=C/C, C\C=C\C, and

Source	Intermediate1	Intermediate2	Target
C=C	CC(C[CH2+])CC	C[CH+]CC	C\C=C\C
C=C	[CH2+]CC		C=CC
C=C	CCC[CH2+]		C=CCC
C=C	CC(C[CH2+])CC	C[CH+]C(C)CC	C\C=C/C

Table 3 Shortest paths for instances where source is set to "C=C" and targets are set to the following: "C\C=C\C", "C=CC", "C=CCC", and "C\C=C/C". Shortest paths are calculated using NetworkX and molecules are listed according to SMILES notation.

Source	Intermediate1	Intermediate2	Target
C=CCC	CCC[CH2+]		C=C
C=CCC	C[CH+]CC		C=CC
C=CCC	C[CH+]CC		C\C=C\C
C=CCC	[CH2+]C(CC)C(C)(C)C	C[CH+]C(C)C(C)(C)C	C\C=C/C

Table 4 Shortest paths for instances where source is set to "C=CCC" and targets are set to the following: "C=C", "C=CC" , "C\C=C\C", "C=CC", and "C\C=C/C". Shortest paths are calculated using NetworkX and molecules are listed according to SMILES notation.

Source	Intermediate1	Intermediate2	Target
C\C=C/C	C[CH+]C(C)C(C)CC	C[CH+]CC	C=CC
C\C=C/C	C[CH+]C(C)C(C)(C)C	[CH2+]C(CC)C(C)(C)C	C=CCC
C\C=C/C	C[CH+]C(C)C(C)(C)C	CC(CC)C([CH2+])(C)C	C=C(C)C
C\C=C/C	C[CH+]C(C)C(C)CC	C[CH+]CC	C\C=C\C

Table 5 Shortest paths for instances where source is set to "C\C=C/C" and targets are set to the following: "C=CC", "C=CCC", "C=C(C)C", and "C\C=C\C". Shortest paths are calculated using NetworkX and molecules are listed according to SMILES notation.

Source	Intermediate1	Intermediate2	Intermediate3	Target
C\C=C\C	C[CH+]CC	CC(C[CH2+])CC		C=C
C\C=C\C	C[CH+]CC			C=CC
C\C=C\C	C[CH+]CC			C=CCC
C\C=C\C	C[CH+]CC	C[C+](C)CC(C)CC		C=C(C)C
C\C=C\C	C[CH+]CC	C[CH+]C(C)C(C)CC		C\C=C/C
C\C=C\C	C[CH+]CC	C=CC	[CH2+]C(C)CCC	C=C(C)CCC
C\C=C\C	C[CH+]CC	CC(C[CH2+])CC	C[CH+]C(C)CC	C\C=C/C

Table 6 Shortest paths for instances where source is set to "C\C=C\C" and targets are set to the following: "C=C", "C=CC", "C=CCC", "C=C(C)C", "C\C=C/C", "C=C(C)CCC", and "C\C=C/CC". Shortest paths are calculated using NetworkX and molecules are listed according to SMILES notation.

Source	Intermediate1	Intermediate2	Target
C=C(C)C	[CH2+]C(C)C		C=CC
C=C(C)C	CC(CC)C([CH2+])(C)C	C[CH+]CC	C=CCC
C=C(C)C	CC(CC)C([CH2+])(C)C	C[CH+]CC	C\C=C\C
C=C(C)C	CC(CC)C([CH2+])(C)C	C[CH+]C(C)C(C)(C)C	C\C=C/C

Table 7 Shortest paths for instances where source is set to "C=C(C)C" and targets are set to the following: "C=CC", "C=CCC", "C\C=C\C", and "C\C=C/C". Shortest paths are calculated using NetworkX and molecules are listed according to SMILES notation.

C=C(C)C, respectively– and target nodes are set to the molecules that are detected experimentally. Note that blanks in Tables 3, 4, 5, 6, and 7 indicate no intermediate in its appropriate column.

By calculating shortest paths, several things can become better understood. To start, C_2H_4, 1-C_4H_8, cis-2-C_4H_8, trans-2-C_4H_8, and iso-C_4H_8– represented by C=C, C=CCC, C\C=C/C, C\C=C\C, and C=C(C)C, respectively– are located in areas that are in more centralized areas of the networks found in Figures 1 and 4. This suggests that these molecules are likely to play important parts in the reactions and intermediate reactions involved in the development of various hydrocarbons. Additionally, the shortest path calculations reveal that there are often not many intermediate reactions that connect the source and target molecules. For example, shortest paths for 1-C_4H_8, as reported in Table 4, mostly involve one intermediate molecule before it connects to its target molecule. Calculating shortest paths also highlights molecules that may have an important role in many different paths. This is seen in the case of trans-2-C_4H_8, where the molecule C[CH+]CC is the first intermediate molecule that appears in every shortest path calculated. C[CH+]CC is also seen to be involved in the shortest paths of other reactions such as the shortest paths leading C=CCC to C=CC and C\C=C\C. Thus, by investigating the shortest paths of detected molecules, one can better understand how molecules may lead to the formation of other molecules and also receive insight on how easily particular reactions may occur.

4 Conclusion

Reaction network of hydrocarbons is investigated via first principle calculations and graph theory. Using networks has allowed for an alternative method of analyzing hydrocarbon formation and decomposition. By transforming hydrocarbon data into a network, it becomes possible to visualize the reactions that influence the development of various hydrocarbons. For instance, one can begin to understand that H+, C=CC, CH_3+, C=C, and [CH_2+]C have heavy influence and has high probability of either forming or reacting with other molecules. Analyzing betweenness centrality has revealed that H+, CH_3+, C_2H_5+, C_8H_{15}+, C_8H_{17}+, and C_6H_{11}+ are found to be molecules that lead to various clusters of additional reactions, suggesting that these molecules are key molecules for additional reactions. Meanwhile, closeness centrality analysis has revealed that H+, CH_3+, C_2H_5+, C_8H_{15}+, and C_8H_{17}+ share high numbers of edges and are thus highly likely to be involved in series of reactions. Network also reveals that H+, C_8H_{17}+, C_8H_{15}+, C_6H_{11}+, and C_6H_{13}+ are very likely to form or react with other molecules, while molecules such as C_8H_{15}+, C_8H_{17}+, and C_6H_{13}+ are also likely to change structures. In addition, expeirmental is performed to validated the results which have a good agreement with calcualted network. Thus, transforming hydrocarbon data into networks provides insight into hydrocarbon cracking and formation, which can be used to refine catalyst design or improving targetted hydrocarbon synthesis.

5 Acknowledgment

The authors wish to thank Itsuki Miyazato for their support in the initial rounds of calculations. This work is funded by Japan Science and Technology Agency(JST) CREST Grant Number JPMJCR17P2.

Notes and references

1 Yang, B.; Sun, W.; Moshammer, K.; Hansen, N. Review of the influence of oxygenated additives on the combustion chemistry of hydrocarbons. *Energy & Fuels* **2021**, *35*, 13550–13568.

2 Chu, S.; Majumdar, A. Opportunities and challenges for a sustainable energy future. *nature* **2012**, *488*, 294–303.

3 Böhm, L. L. The ethylene polymerization with Ziegler catalysts: fifty years after the discovery. *Angewandte Chemie International Edition* **2003**, *42*, 5010–5030.

4 Weitkamp, J. Catalytic hydrocracking—mechanisms and versatility of the process. *ChemCatChem* **2012**, *4*, 292–306.

5 McDaniel, M. P. A review of the Phillips supported chromium catalyst and its commercial use for ethylene polymerization. *Advances in catalysis* **2010**, *53*, 123–606.

6 Torres Galvis, H. M.; de Jong, K. P. Catalysts for production of lower olefins from synthesis gas: a review. *ACS catalysis* **2013**, *3*, 2130–2149.

7 Schwach, P.; Pan, X.; Bao, X. Direct conversion of methane to value-added chemicals over heterogeneous catalysts: challenges and prospects. *Chemical reviews* **2017**, *117*, 8497–8520.

8 Farrauto, R. J.; Hobson, M.; Kennelly, T.; Waterman, E. Catalytic chemistry of supported palladium for combustion of methane. *Applied Catalysis A: General* **1992**, *81*, 227–237.

9 Choudhary, T.; Banerjee, S.; Choudhary, V. Catalysts for combustion of methane and lower alkanes. *Applied Catalysis A: General* **2002**, *234*, 1–23.

10 Elvers, B., et al. *Ullmann's encyclopedia of industrial chemistry*; Verlag Chemie Hoboken, NJ, 1991; Vol. 17.

11 Li, Y.; Qi, F. Recent applications of synchrotron VUV photoionization mass spectrometry: insight into combustion chemistry. *Accounts of chemical research* **2010**, *43*, 68–78.

12 Bai, J.; Liu, X.; Lei, T.; Teng, B.; Wen, X. A combined DFTB nanoreactor and reaction network generator approach for the mechanism of hydrocarbon combustion. *Chemical Communications* **2021**, *57*, 11633–11636.

13 Wolfgang, R. Energy and chemical reaction. ii. intermediate complexes vs. direct mechanisms. *Accounts of Chemical Research* **1970**, *3*, 48–54.

14 Takahashi, L.; Ohyama, J.; Nishimura, S.; Takahashi, K. Representing the Methane Oxidation Reaction via Linking First-Principles Calculations and Experiment with Graph Theory. *The Journal of Physical Chemistry Letters* **2020**, *12*, 558–568.

15 Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nature communications* **2017**, *8*, 1–7.

16 Mutlay, I.; Restrepo, A. Complex reaction networks in high temperature hydrocarbon chemistry. *Physical Chemistry Chemical Physics* **2015**, *17*, 7972–7985.

17 Jacox, M. E. The spectroscopy of molecular reaction intermediates trapped in the solid rare gases. *Chemical Society Reviews* **2002**, *31*, 108–115.

18 Roithová, J. Characterization of reaction intermediates by ion spectroscopy. *Chemical Society Reviews* **2012**, *41*, 547–559.

19 Dong, J.-C.; Zhang, X.-G.; Briega-Martos, V.; Jin, X.; Yang, J.; Chen, S.; Yang, Z.-L.; Wu, D.-Y.; Feliu, J. M.; Williams, C. T., et al. In situ Raman spectroscopic evidence for oxygen reduction reaction intermediates at platinum single-crystal surfaces. *Nature Energy* **2019**, *4*, 60–67.

20 Grambow, C. A.; Jamal, A.; Li, Y.-P.; Green, W. H.; Zador, J.; Suleimanov, Y. V. Unimolecular reaction pathways of a γ-ketohydroperoxide from combined application of automated reaction discovery methods. *Journal of the American Chemical Society* **2018**, *140*, 1035–1048.

21 Suleimanov, Y. V.; Green, W. H. Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods. *Journal of chemical theory and computation* **2015**, *11*, 4248–4259.

22 Zimmerman, P. M. Single-ended transition state finding with the growing string method. *Journal of computational chemistry* **2015**, *36*, 601–611.

23 Van de Vijver, R.; Zádor, J. KinBot: Automated stationary point search on potential energy surfaces. *Computer Physics Communications* **2020**, *248*, 106947.

24 O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of cheminformatics* **2011**, *3*, 1–14.

25 Landrum, G., et al. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. 2013.

26 Bastian, M.; Heymann, S.; Jacomy, M. Gephi: an open source software for exploring and manipulating networks. Third international AAAI conference on weblogs and social media. 2009.

27 Hagberg, A.; Swart, P.; S Chult, D. *Exploring network structure, dynamics, and function using NetworkX*; 2008.