

tsscds2018: A Code for Automated Discovery of Chemical Reaction Mechanisms and Solving the Kinetics

Aurelio Rodríguez,^[a] Roberto Rodríguez-Fernández,^[b] Saulo A. Vázquez,^[b]
George L. Barnes,^[c] James J. P. Stewart^[d] and Emilio Martínez-Núñez^{*,[b]}

A new software, called tsscds2018, has been developed to discover reaction mechanisms and solve the kinetics in a fully automated fashion. The program employs algorithms based on Graph Theory to find transition state (TS) geometries from accelerated semiempirical dynamics simulations carried out with MOPAC2016. Then, the TSs are connected to the corresponding minima and the reaction network is obtained. Kinetic data like populations vs time or the abundancies of each

product can also be obtained with our program thanks to a Kinetic Monte Carlo routine. Highly accurate *ab initio* potential energy diagrams and kinetics can also be obtained using an interface with Gaussian09. The source code is available on the following site: <http://forge.cesga.es/wiki/g/tsscds/HomePage>
© 2018 Wiley Periodicals, Inc.

DOI:10.1002/jcc.25370

Introduction

The study of mechanisms of chemical reactions can greatly benefit from the use of computational methods, either because of the experimental difficulties to characterize fleeting intermediates, or just because they nicely supplement spectroscopic methods and/or experimental studies of isotope effects.^[1–4]

Although the strategy commonly adopted in most computational studies aimed to explain reaction mechanisms is to employ chemical intuition, a growing number of automated computational techniques have flourished in the last years.^[5–57] The emergence of automated computational protocols has helped elucidate complex reaction networks and minimized the risk of overlooking meaningful mechanisms in manual searches.

We have recently shown that the use of accelerated direct dynamics in combination with an efficient geometry-based algorithm to identify bond breaking/formation can be very convenient tools to find transition states (TSs) of chemical reactions.^[49–51] The novelty of our method, termed transition state search using chemical dynamics simulations (*tsscds*), is the direct search for saddle points from the simulation snapshots. Since the accelerated dynamics is non-IRC,^[58] our procedure provides TS structures and pathways not only for unimolecular processes but also for bimolecular ones, which makes it very powerful. Besides the discovery of reaction mechanisms, the kinetics is solved using Kinetic Monte Carlo,^[59] thus providing useful information of product yields.^[50]

In this manuscript, we present the computer program named *tsscds2018*, which performs all the tasks mentioned in the previous paragraph, including some improvements over the original procedure, as well as additional tools. For instance, the initial energy/temperature for the simulations is determined automatically from the size of the system (being proportional to the number of degrees of freedom) and RRK theory.^[60] Besides initializing the dynamics simulations from multiple minima as described previously,^[50] a range of initial energies/

temperatures is employed to minimize a bias in the TS search toward high-entropy structures. Finally, different biased dynamics methods are now available in the program, which greatly enhance its efficiency compared to the original procedure, where the chemical reactions were solely driven by the use of high energies or temperatures.

The structure of *tsscds2018* is modular so that new capabilities can be easily added. The accelerated dynamics is carried out by DRC module of MOPAC2016,^[61] which has been modified to incorporate the aforementioned biased dynamics methods. In the first instance, the reaction mechanisms and kinetics are thus obtained using a semiempirical level of theory. However, more accurate results can be obtained by re-optimizing all stationary points at an *ab initio*/DFT level of theory using Gaussian09.^[62] A handy feature is the use of GNU Parallel,^[63] allowing parallel executions of different modules, which results in an overall enhancement of the performance. To facilitate and simplify management of the structures/input files, SQLite3 databases are employed. Finally, some scripts

[a] A. Rodríguez

Galicia Supercomputing Center (CESGA), Santiago de Compostela, Spain

[b] R. Rodríguez-Fernández, S. A. Vázquez, E. Martínez-Núñez

Departamento de Química Física, Facultade de Química, Campus Vida, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain

E-mail: emilio.nunez@usc.es

[c] G. L. Barnes

Department of Chemistry and Biochemistry, Siena College, 515 Loudon Road, Loudonville, New York

[d] J. J. P. Stewart

Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs, Colorado, 80921

Grant sponsor: Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia; Grant numbers: ED431C 2017/17 and CTQ2014-R; Grant sponsor: The National Institute of General Medical Sciences of the National Institutes of Health; Grant number: R44GM108085; Grant sponsor: National Science; Grant numbers: CHE-1763652

© 2018 Wiley Periodicals, Inc.

employ dialog boxes to input data and progress bars to monitor the calculations, which are implemented using Zenity.

The original *tsscds* method is very briefly described first, then the structure of *tsscds2018* is presented. The use of the program is most conveniently explained through a simple test case: the unimolecular decomposition of formic acid (FA), which is detailed in a separate section. Further capabilities are outlined next, and different applications carried out in our lab are also described. Finally, a summary and future improvement are discussed in the last section.

Methods

The method, also named *tsscds*, has been recently developed by one of the authors,^[50,51] and it has been devised to find transition states or, more precisely, first-order saddle points in a molecular system. The basic idea behind *tsscds* is to run accelerated (high-temperature or high-energy) semiempirical direct dynamics simulations to break/form new bonds within the first few hundred femtoseconds. Then, an efficient post-processing algorithm identifies geometries with partly formed/broken bonds, which serve as guess structures for transition state optimizations. Once the TSs are optimized, a reaction network can be constructed by computing the intrinsic reaction coordinates (IRCs),^[64] which connect TSs with minima. The method employs two levels of theory: semiempirical and *ab initio*/DFT. The semiempirical calculations are performed to run the direct dynamics and to obtain approximate TSs structures, while a higher level of theory is used to re-optimize the TSs and run IRC calculations. Two different electronic structure programs are employed: MOPAC2016^[61] and Gaussian 09^[62] for the semiempirical and *ab initio*/DFT calculations, respectively.

Once the fully connected stationary points are obtained, rate coefficients for each elementary step are calculated from statistical theories,^[50,51] and the kinetics are solved using Kinetic Monte Carlo (KMC).^[59]

Structure of the program

The program finds reaction pathways and solves the kinetics at two levels of theory, as aforementioned. Two scripts, **llcalcs.sh** and **hcalcs.sh**, have been written to carry out all the low-level (ll) and high-level (hl) calculations, respectively. Each of them, in turn, is made up of different modules/programs (written in Bash shell scripting, Python2 and Fortran 90) to carry out specific tasks.

Figure 1 shows a flowchart of **llcalcs.sh**. As seen in the figure, the script has several components. It starts with a loop, shown on the left, that will be carried out for a given number of cycles or iterations (*niter*). The loop starts executing **tsscds_parallel.sh**, which submits a number of parallel and independent accelerated dynamics simulations (*ntasks*) using MOPAC2016.^[61] Figure 2 shows an example of a **tsscds_parallel.sh** job consisting of 4 parallel tasks, each of them performed by a script called **tsscds.sh**. In turn, the first step of **tsscds.sh** consists of selecting initial Cartesian coordinates **q** and momenta **p** using a microcanonical or a canonical ensemble,^[65,66] which is done by

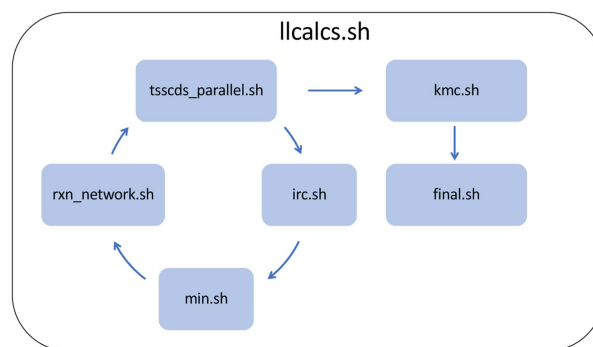


Figure 1. Flowchart of **llcalcs.sh** script. [Color figure can be viewed at wileyonlinelibrary.com]

either **nm.exe** or **termo.exe**, respectively. Both programs are written in fortran90. The initial energy or temperature of the system is chosen automatically by the program (the reader is referred to the tutorial^[67] for details). Having **q** and **p** been chosen, **tsscds.sh** runs now a number of trajectories (*ntraj*) using a locally modified version of DRC module in MOPAC2016. Details of the modified DRC module are given in the tutorial. Once the accelerated dynamics calculations are completed, **bbfs.exe** (written in fortran90) locates guess TS structures from the geometries along the trajectories.^[51] Finally, MOPAC2016 optimizes the transition states using the standard Eigenvector Following algorithm.^[68]

After **tsscds_parallel.sh** has completed all (parallel) tasks, **irc.sh** screens the obtained structures to remove possible redundancies and/or saddle points associated with van der Waals intermediates (see Fig. 1). Following completion of the screening, IRC calculations are carried out in both the forward and backward directions.^[64]

The last points of each IRC are the initial guesses of subsequent optimizations carried out by **min.sh**, a procedure whereby each TS is connected with the corresponding minimum energy structures. Thus, a reaction network is built, and each structure is labeled as either intermediate, or product (containing several fragments). In addition, groups of conformational isomers are identified, which is very useful to carry out coarse-grained kinetics simulations as discussed below. The construction of the reaction network and labeling of the different structures is performed by **rxn_network.sh** script.

As seen in Figure 1, **rxn_network.sh** closes the loop, and its output is fed into **tsscds_parallel.sh**. In particular, the newly generated minima are needed by **tsscds_parallel.sh** because the ensembles of trajectories are initialized not only from the starting structure but also from the new minima. When a maximum number of iterations is reached, the kinetics is solved using **kmc.sh** (see Fig. 1), which performs Kinetic Monte Carlo^[59] (KMC) simulations. This script calculates rate coefficients for every single elementary step and employs a fortran90 KMC program to obtain, as a function of time, the populations of all the chemical species involved in the reaction network. Finally, **final.sh** gathers all relevant mechanistic and kinetics information obtained throughout the calculations.

As aforementioned, the reaction network and kinetic results can also be obtained using an *ab initio*/DFT level of theory with

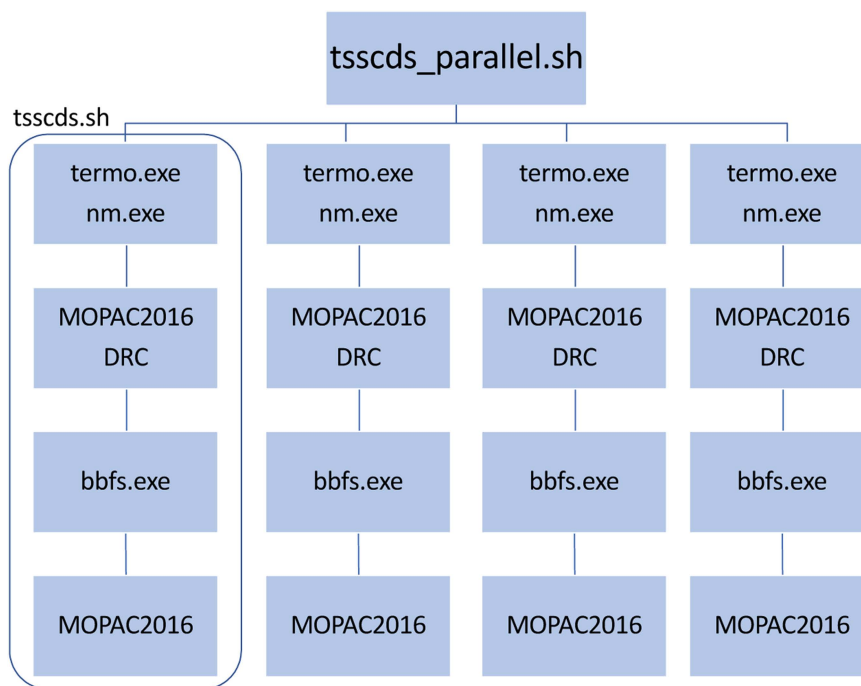


Figure 2. Different tasks carried out by **tsscds_parallel.sh**. In this example a total of 4 independent **tsscds.sh** jobs are carried out in parallel. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

G09. The high-level tasks are performed with **hlcals.sh**, which is the counterpart of **lcalcs.sh** described in the previous paragraphs. Since the TSs have already been found at low-level, the low-level-optimized structures are now the initial guesses for the high-level optimization. In addition, the product fragments are now optimized to construct more accurate potential energy diagrams. Therefore, the structure of **hlcals.sh** is somehow different from that of **lcalcs.sh**, as seen in Figure 3. Specifically, the different tasks carried out by each component of **hlcals.sh** are 1) high-level optimization of the TSs obtained at low-level (**TS.sh**); 2) high-level IRC calculations from the TSs optimized in the previous step (**IRC.sh**); 3) high-level optimization of the corresponding intermediates (**MIN.sh**); 4) construction of the high-level network (**RXN_NETWORK.sh**); 5) kinetics simulations on the high-level network (**KMC.sh**); 6) high-level optimization of the products (**PRODs.sh**) and 7) Gathering of the important mechanistic and kinetic results (**FINAL.sh**).

Discovering the reaction mechanisms of formic acid

This section deals with the use of the program to study a simple example. Specifically, we have chosen the dissociation of formic acid (FA) as a test case. The reader is referred to the tutorial that comes with this distribution^[67] for detailed instructions to install the program as well as for a thorough explanation of the program execution and input/output files.

Description of the input files. Only two input files (**FA.xyz** and **FA.dat**) are needed to run this example. The first file, **FA.xyz**, where FA is the name of the system, contains the Cartesian coordinates of the system, usually the most stable conformer of the reactant molecule. The second file, **FA.dat**, contains all

parameters of the calculation. Figure 4 shows an example of this input file. As can be seen, the file is split in 5 different sections. Each line, within each of the sections, starts with a (case sensitive) keyword, followed by some values or arguments.

In the General section, the user provides keywords and their arguments or parameters necessary for the electronic structure calculations. Specifically, *molecule* refers to the name of the system and *mult* to its multiplicity (1 in the present case). The keyword *LowLevel* is used to specify the semiempirical method employed to run the trajectories (conducted by MOPAC2016). If the selected method is the default (PM7), this keyword is unnecessary. *HighLevel* determines the level of theory employed in the high-level calculations. *HL_rxn_network* indicates whether the high-level reaction network is calculated starting from all the obtained low-level TSs (and the keyword should be followed by *complete*), or whether bimolecular reactions are removed, in which case *reduced* should be employed instead.

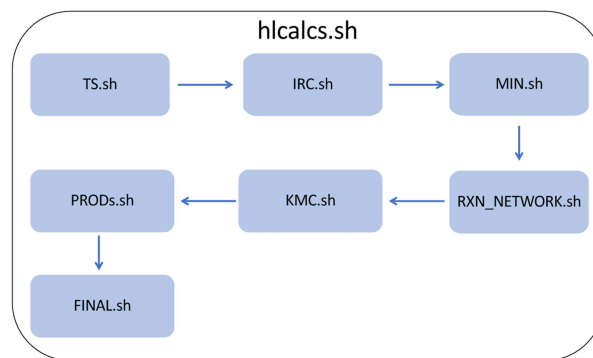


Figure 3. Flowchart of **hlcals.sh** script. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

```
--General section--
molecule FA
HighLevel b3lyp/6-31G(d,p)
HL_rxn_network complete
charge 0
mult 1

--CDS section--
sampling microcanonical
ntraj 10

--BBFS section--
freqmin 200

--Screening of the structures section--
avgerr 0.008
bigerr 2.5
thdiss 0.1

--Kinetics section--
Rate microcanonical
EKMC 150
```

Figure 4. Input file *FA.dat* employed to study the decomposition of formic acid.

The next section is called CDS (for Chemical Dynamics Simulations). Here, the user provides details of the accelerated dynamics simulations. In this example, we employ *microcanonical sampling* to select the initial **q** and **p** for the accelerated semiempirical dynamics simulations. Other *sampling* options are explained in the tutorial. The keyword *ntraj* refers to the number of trajectories.

The BBFS (Bond Breaking/Formation Search) section deals with the criteria used for the selection of structures from the trajectory results.^[51] In the present example, using keyword *freqmin*, TS structures with imaginary frequencies lower than 200 cm⁻¹ will be disregarded. This keyword is related to the next section.

In the section "Screening of the structures," *tsscds* screens the obtained TS structures for possible redundancies. Furthermore, some TSs may correspond to floppy van der Waals complexes formed upon fragmentation, which might not be important in many systems. To avoid or minimize repeated structures and van der Waals complexes, the program includes a screening tool that employs Spectral Graph Theory to calculate the following quantities: SPRINT coordinates,^[69] degrees of each vertex and eigenvalues of the Laplacian matrix.^[50] Comparing these values (including the energy) for two structures, the mean absolute percentage error (MAPE) and the biggest absolute percentage error (BAPE) are obtained. The keywords *avgerr* and *bigerr* refer to the maximum values for MAPE and BAPE, respectively. If both the MAPE and BAPE values calculated for two structures are below *avgerr* and *bigerr*, respectively, the structures are regarded as equal.

The last keyword, called *thdiss*, refers to the eigenvalues of the Laplacian (EL). In Spectral Graph Theory, the number of eigenvalues equal to zero provides the number of connected graphs, which is translated here as the number of fragments in the molecular system. The keyword *thdiss* refers to the threshold for an EL to be considered 0. For instance, in our example, if an EL < 0.1 (see Fig. 4), then, this EL is set to 0. This keyword

is used to identify van der Waals complexes that are formed upon unimolecular fragmentation.

In the Kinetics section, the user provides details for the kinetics calculations that simulate the experimental conditions. The keyword *Rate* has one of the following arguments: *canonical* or *microcanonical*. These arguments tell the program to compute rate constants according to Transition State Theory (TST) or Rice–Ramsperger–Kassel–Marcus (RRKM) theory, respectively. Then, a keyword called *EKMC* or *TKMC* is used to specify the excitation energy (in kcal/mol) or the temperature for the calculation of rate constants for microcanonical or thermal (canonical) systems, respectively. At present, temperatures below 100 K are not allowed.

Brief description of the main output files. As aforementioned, **final.sh** (or its high-level counterpart **FINAL.sh**) gathers all relevant information in a folder named FINAL_XL_FA (where XL = HL, LL for high-level and low-level, respectively). These folders contain several files that summarize the most relevant results. The most important files are described as follows (for details see the tutorial).

Files *MINinfo* and *TSinfo* list the located minima and TSs, respectively, together with the corresponding relative energies. In these files, identification integers are used independently for minima and transition states.

Each FINAL_XL_FA (XL = HL or LL) folder includes three SQLite3 tables (having the *db* extension) that contain the geometries, energies and frequencies of minima, products and TSs. The user can easily extract information from these tables using the **select.sh** script, as specified in the tutorial.

A file called *RXNet* contains information of the reaction network. For each TS, the file specifies the associated minima and/or products and their corresponding identification numbers. Also, the chemical formulas of the product fragments are listed at the end of the file. File *RXNet.cg* is similar to *RXNet*, but gives information of the coarse-grained KMC calculations, which is the default approach in *tsscds2018*. In this approach, conformational isomers form a single species, which is taken as the lowest energy isomer.^[50] The FINAL_XL_FA (XL = HL, LL) folders also contain a file, called *RXNet.rel*, which lists only the kinetically-relevant channels, that is, those that intervene in at least 0.1% of the total number of processes.

File *kineticsFvalue* contains the kinetic results, namely, the final branching ratios and the population of every species as a function of time. In the actual filename, "F" is either "T" or "E," depending on whether the simulations were carried out for a given temperature or energy, respectively, and "value" is the corresponding (temperature or energy) value. For instance, the kinetic results for a canonical calculation at 298 K would be printed in a file called *kineticsT298*.

There are two files, called *Energy_profile.gnu* and *populationFvalue.gnu*, which contain data for gnuplot. The former can be used to plot an energy diagram of the kinetically-relevant paths obtained at the simulated conditions. The latter file provides the population of each species as a function of time. "F" and "value" in *populationFvalue.gnu* follow the same rules described above.

Finally, each of the FINAL_XL_FA folders includes a directory named *normal_modes*, which contains the normal mode eigenvectors and eigenvalues of TSs and minima. They are specified in Molden format, for visualization with this graphic software.

Details of the kinetics simulations. As indicated above, by default, the KMC simulations regard conformational isomers as a single species, which speeds up the calculations.^[50] However, each conformational isomer could be treated as a single species in the KMC calculations. If that is the case, the reaction network needs to be reconstructed and the kinetics simulations carried out on the extended network. That entails, for the low-level calculations, running again **rxn_network.sh** (using *allstates* as argument), **kmc.sh** and **final.sh**, as described in the tutorial. The corresponding procedure for the high-level calculations is similar, except for the use of the scripts in capital letters (**RXN_NETWORK.sh**, **KMC.sh** and **FINAL.sh**).

In addition, when the calculations seek to simulate a thermal experiment (i.e., when *rate canonical* is employed in the input file), the kinetic results can be rerun for a different temperature from that specified in the input file through the keyword *TKMC*. This can be easily done using the **kinetics.sh** with the appropriate arguments (see the tutorial^[67] for details).

Other capabilities of the code

Besides the above basic features, additional tools are implemented in the program. The reader is referred to the tutorial for a thorough explanation of the other capabilities, as here only a very brief summary is provided. The additional tools of the program involve other sampling options (*intermolecular* and *external*) and further accelerated dynamics choices.

The options *intermolecular* and *external* are employed to optimize intermolecular complexes, and to use external programs to carry out the dynamics simulations, respectively. The latter option has been added to interface the chemical dynamics simulation code VENUS^[70] with *tsscds2018*. This feature could be of great interest to simulate mass spectrometry experiments, where collisions with projectiles are employed to dissociate the molecule.^[71]

Finally, other accelerated dynamics techniques have been included in the program, like the use of phase space constraints^[72] or bias potentials. Examples of how to use those features with some simple examples are given in the tutorial.

Applications

The study of the decomposition of formic acid was presented in the previous section, and in more detail in the tutorial, as an example for the illustration of the main features of *tsscds2018*. The program, however, has been employed in our lab to elucidate reaction mechanisms in different types of systems. In this section, we summarize the most relevant results.

Small systems

The smallest systems studied with our procedure are formaldehyde, formic acid (FA) and vinyl cyanide (VC),^[51] for which a

total of 7, 12 and 83 TS structures, respectively, were located with the first version of *tsscds*. Of significance, a new TS for the water-gas shift reaction (WGS: $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2$) was found for FA.^[51] This is an interesting result since the WGS is bimolecular, whereas the accelerated dynamics is unimolecular, which exemplifies the highly non-IRC nature of our simulations and the wealth of information (structures) that can be drawn using our methodology. Also, the theoretical VC decomposition kinetics, studied separately,^[73] led to nearly perfect agreement with the experimental HCN/HNC branching ratio.

Medium-size systems

The fragmentation mechanisms of medium-size systems in the gas phase are, in general, very complex. For example, the dissociation of propenal comprises many different fragmentation channels that involve well over 250 transition states.^[50] The branching ratios calculated with the *tsscds* program for the different dissociation channels of the molecule agree very well with the available experimental data.

Three novel HCl dissociation channels of acryloyl chloride (AC), which had gone unnoticed in previous theoretical work, were very recently discovered with our procedure.^[74] They are displayed in red in Figure 5. Two of these elementary reactions are three-body dissociations leading to acetylene, carbon monoxide and hydrogen chloride (left of the figure), and they become important at high excitation energies. And the other new five-center mechanism (leading to $\text{HCCHCO} + \text{HCl}$ in the right side of the figure) is the predominant at the experimental conditions.

The collision-induced dissociation (CID) of protonated uracil is very complex. Using *tsscds*, we discovered more than one thousand stationary points and 751 elementary reactions.^[71] Figure 6 depicts the 14 lowest-energy structures of protonated uracil found with our automated code at the B3LYP/6-31+G(d,p). For simplicity, only the lowest-energy conformer of each isomer is displayed. Of the total isomers shown in the figure, only the first, second, fourth, eighth, tenth and fourteenth structures were considered in a previous work on this system, in which the

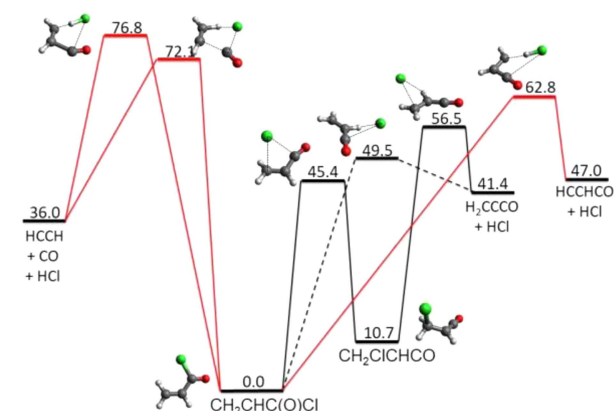


Figure 5. Schematic diagram of the most important HCl elimination reactions on the ground electronic state of acryloyl chloride. Relative energies are given in kcal mol^{-1} . [Color figure can be viewed at wileyonlinelibrary.com]

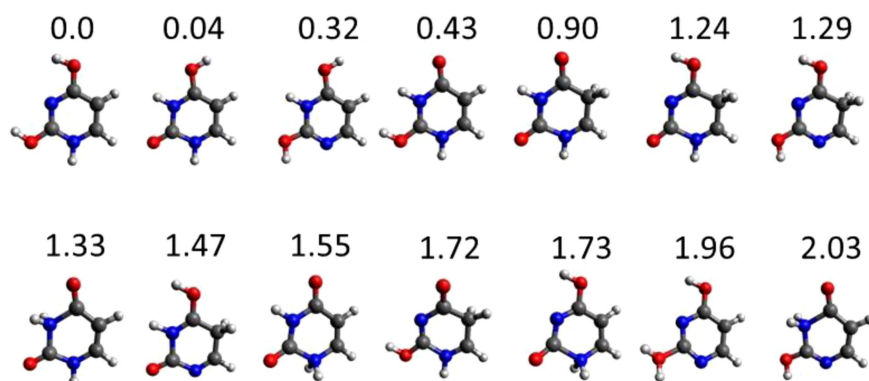


Figure 6. Lowest-energy isomers of protonated uracil. Relative energies (in kcal mol⁻¹) were obtained by B3LYP/6-311+G(3df,2p) single point calculations using B3LYP/6-31+G(d,p) geometries. [Color figure can be viewed at wileyonlinelibrary.com]

potential energy surface (PES) was explored by hand and based on chemical intuition.

Relative abundances of fragmentation products, as a function of the ion's internal energy, were determined by KMC simulations, using microcanonical rate constants calculated by Rice–Ramsperger–Kassel–Marcus (RRKM) theory. It is possible to compare these abundances with those determined in the CID experiments, as shown in Figure 7, using a correspondence between the collision energy in the center-of-mass framework (E_{com}) and the ion's internal energy, obtained by dynamics simulations.^[71] As can be seen, for the predominant dissociation channels, the product abundances calculated by our code are in qualitative agreement with experiment, taking into account the complexity of the system.

In one of our latest applications, we employed the program to explore possible sources of HCN and HNC formation in astrophysical environments.^[75] In particular, time-resolved infrared spectroscopy experiments detected formation of both HCN and HNC after 193-nm photolysis of methyl cyanoformate. Our automated protocol was able to locate several cyclic TSs leading to HNC and HCN on the ground-state potential energy surface, as shown in Figure 8. The HCN/HNC branching ratio obtained in our simulations (0.01) is in semiquantitative agreement with that determined in the experiments (≈ 0.07). Furthermore, quasi-classical trajectory deduced internal energy distributions of HCN and HNC are in very good agreement with the experimental ones, which supports that

formation of HCN and HNC in the photodissociation of methyl cyanoformate takes place on the ground-potential energy surface, after internal conversion from the initial excited states. The work provides further insights into the intriguing observation of overabundance of HCN in astrophysical environments.

Reactants, intermediates and transition states in medium and large systems commonly have several conformers or rotamers. For example, let us consider the thermal decomposition of 1-propanol radicals. The most relevant pathways reported in the literature^[76–82] are displayed in Figure 9. We applied our automated code to this system^[83] and we obtained all the channels shown in the figure, except the barrierless dissociation leading to propene + OH, since the present version of tsscds2018 cannot handle this type of reactions. Importantly, we found a wealth of reactant and TS conformers, not described in the previous studies, as indicated by the numbers of rotamers shown in the figure (they do not include enantiomers). For convenience, we can define a conformational reaction channel (CRC) as the group of all the paths, including specular images, that connect the conformers of a given reactant with the corresponding TS conformers. This definition facilitates the analyses of the kinetics of complex reaction networks.

The presence of several rotamers may have a significant influence on the rate constants and branching ratios. Using variational transition state theory (VTST),^[84–86] we computed rate

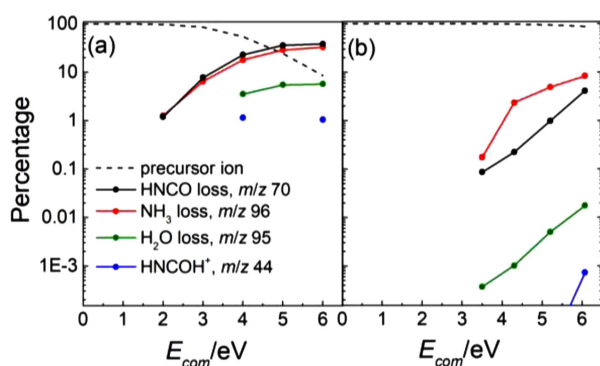


Figure 7. a) Experimental and b) calculated intensities of precursor and fragment ions produced by CID of protonated uracil. [Color figure can be viewed at wileyonlinelibrary.com]

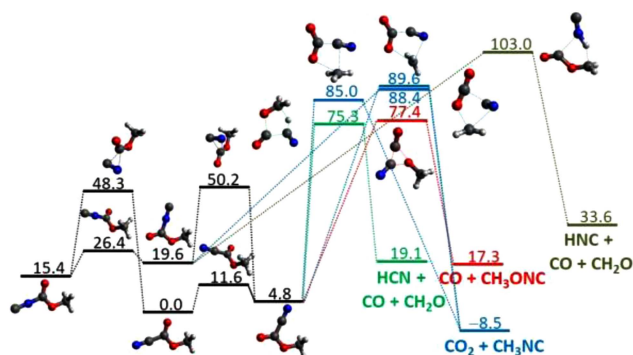


Figure 8. Relevant decomposition pathways in the ground-state PES of methyl cyanoformate. Relative energies (in kcal mol⁻¹) include ZPE contributions and were obtained by CCSD(T)/6-311++G(3df,3pd)//MP2/6-311+G(2d,2p) calculations. [Color figure can be viewed at wileyonlinelibrary.com]

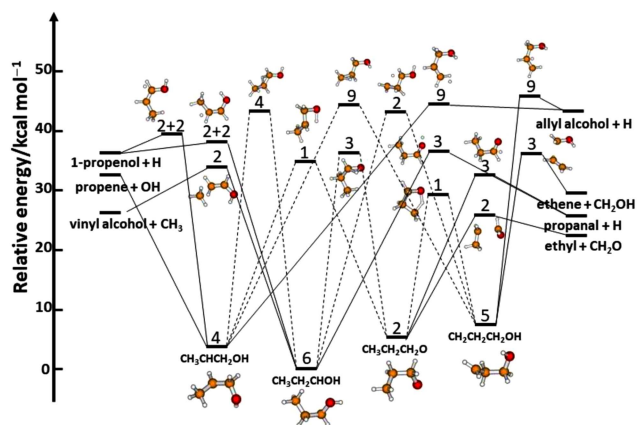


Figure 9. Relevant CRCs in the thermal decomposition of 1-propanol radicals. The digits give the numbers of conformers for reactants and TSs. 2 + 2 refers to *cis* and *trans* conformations (2 of each type). [Color figure can be viewed at wileyonlinelibrary.com]

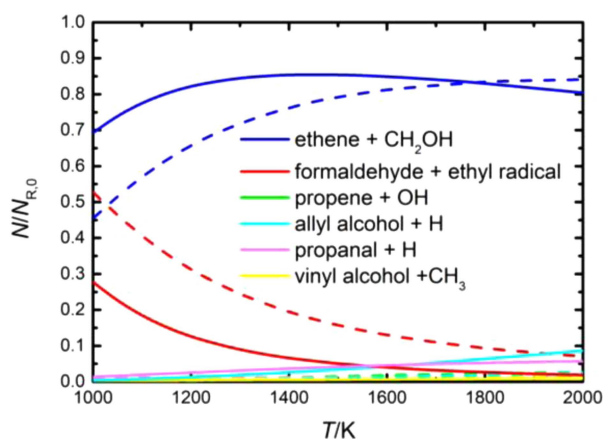


Figure 10. Product branching ratios obtained from KMC simulations taken the $\text{CH}_2\text{CH}_2\text{CH}_2\text{OH}$ radical as the reactant. Solid and dashed lines correspond to the MP and 1W simulations, respectively. $N_{R,0}$ is the initial number of $\text{CH}_2\text{CH}_2\text{CH}_2\text{OH}$ radicals and N is the number of a given product. [Color figure can be viewed at wileyonlinelibrary.com]

constants for all the CRCs depicted in Figure 9. In particular, we used the multipath approach (MP-VTST),^[86–90] with which the rate constant of a given CRC is calculated using contributions

from all the conformers and paths, as well as the simplest one-well (1W) approach, in which only the most stable conformer of the reactant and that of the TS are considered. The latter is the common approach used in most studies. We found important discrepancies between the rate constants calculated by these two approaches.^[83] In addition, KMC simulations using the MP and 1W sets of rate constants led to significant deviations in product abundances, as depicted in Figure 10 for kinetics started at the $\text{CH}_2\text{CH}_2\text{CH}_2\text{OH}$ radical.

These results show the importance of considering all the conformers and paths, which, in turn, reinforces the need of using automated codes for discovering reaction mechanisms due to the extra complexity of the PESSs. As indicated previously, the *tsscds* user can construct the reaction network with all the paths or, alternatively, with CRCs for which TST rate constants are calculated by taken into account the number of reactant and TS conformers (coarse-grained approach), as detailed elsewhere.^[50]

Organometallic catalysis

A step further in the complexity of reaction mechanisms appears in the field of organometallic catalysis, in which our program was successfully applied. In particular, the cobalt-catalyzed hydroformylation of ethylene was the chosen test case.^[49]

The study entailed running *tsscds* in eight different systems, which involved combinations of the starting materials (CO , H_2 and ethylene) with the catalyst $\text{Co}(\text{CO})_3$. After merging all results, the kinetics simulations give rise to a theoretical rate law for hydroformylation that agrees rather well with the experimental one and with that obtained from highly accurate *ab initio* calculations^[91] (see Fig. 11). In addition, our method predicts that hydrogenation of ethylene is a side reaction that can be predominant under certain experimental conditions.

Conclusions

A user-friendly program for the discovery of reaction mechanisms and for efficiently solving the kinetics is presented in this manuscript. The code relies on exploratory semiempirical accelerated dynamics simulations carried out by a modified version

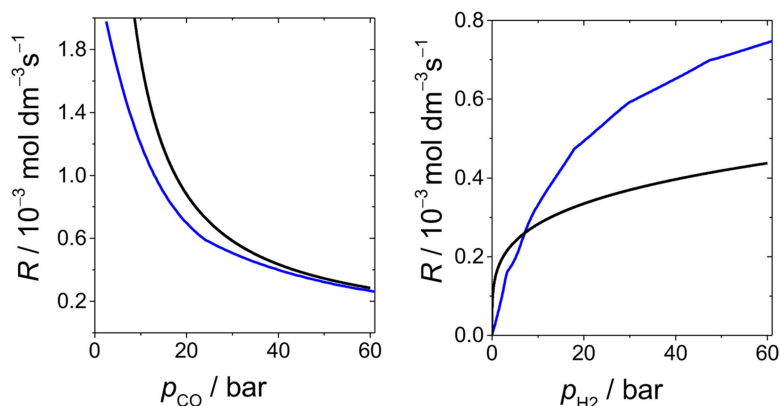


Figure 11. Rates of hydroformylation as a function of CO and H_2 pressures obtained with *tsscds* (black) compared with the rates obtained from the highly accurate *ab initio* calculations of Harvey and co-workers (blue).^[91] [Color figure can be viewed at wileyonlinelibrary.com]

of MOPAC2016, and on an efficient geometry-based algorithm recently developed by one of the authors. The resulting potential energy diagrams and kinetics can be obtained not only at the semiempirical level, but also at higher (*ab initio*/DFT) levels using G09.

The only input needed from the user is a file containing some details of the calculations, and an initial input structure, that can be taken from experiments, previous computations, or even constructed with any visualization graphics software. The procedure is therefore fully automated, except for the selection of three screening parameters that serve to avoid redundant structures and TSs connecting van der Waals minima. In that case, the user is advised to check a few structures with the naked eye to judge the validity of the input screening parameters. After that, several components of the program undertake the different tasks needed to generate potential energy diagrams and plots of populations vs time.

Several new tools will be incorporated in a future release, and our team has already started working on the following features:

Treatment of gas-phase barrierless reactions. BBFS algorithm is responsible for the identification of guess TSs that will be then subjected to the standard EF algorithm to optimize saddle points. Therefore, processes that occur without a potential-energy barrier are elusive, and oftentimes they are predominant. Identification of those processes is not a major issue, but an accurate automated evaluation of the involved rate coefficients is not a trivial task.

A second tool that will be incorporated in future versions is the analysis of secondary fragmentations. That entails running the accelerated dynamics not only from the input molecule but also from the fragments that result upon dissociation of the parent molecule. This feature is very important, for instance, in the theoretical analysis of mass spectrometry experiments.

The third feature that will be present in a future release is the analysis of bimolecular reactions. Even though one can already start from any given structure of the system, including shallow van der Waals minima, it would be desirable to start the dynamics from the separated chemical species and run the bimolecular dynamics. That entails the implementation of appropriate samplings of initial conditions and an ample selection of the initial relative velocities and energies of each fragment.

Acknowledgment

Authors gratefully thank CESGA (Galicia Supercomputing Center, Santiago de Compostela, Spain) for providing access to computing facilities.

Keywords: accelerated dynamics simulations · Graph Theory · reaction network · Kinetic Monte Carlo · automated method

How to cite this article: A. Rodríguez, R. Rodríguez-Fernández, S. A. Vázquez, G. L. Barnes, J. J. P. Stewart, E. Martínez-Núñez. *J. Comput. Chem.* **2018**, 39, 1922–1930. DOI: 10.1002/jcc.25370

- [1] Q. Peng, R. S. Paton, *Acc. Chem. Res.* **2016**, 49, 1042.
- [2] T. Sperger, I. A. Sanhueza, F. Schoenebeck, *Acc. Chem. Res.* **2016**, 49, 1311.
- [3] T. Sperger, I. A. Sanhueza, I. Kalvet, F. Schoenebeck, *Chem. Rev.* **2015**, 115, 9532.
- [4] H. Schwarz, *Angew. Chem. Int. Ed.* **2011**, 50, 10096.
- [5] H. B. Schlegel, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, 1, 790.
- [6] H. L. Davis, D. J. Wales, R. S. Berry, *J. Chem. Phys.* **1990**, 92, 4308.
- [7] J. Q. Sun, K. Ruedenberg, *J. Chem. Phys.* **1993**, 98, 9707.
- [8] C. J. Tsai, K. D. Jordan, *J. Phys. Chem.* **1993**, 97, 11227.
- [9] Y. Abashkin, N. Russo, *J. Chem. Phys.* **1994**, 100, 4477.
- [10] K. Bondensgard, F. Jensen, *J. Chem. Phys.* **1996**, 104, 8025.
- [11] J. P. K. Doye, D. Wales, J. Z. Phys. D Atoms Mol. Clusters **1997**, 40, 194.
- [12] W. Quapp, M. Hirsch, O. Imig, D. Heidrich, *J. Comput. Chem.* **1998**, 19, 1087.
- [13] M. Černohorský, S. Kettou, J. Koča, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 705.
- [14] K. M. Westerberg, C. A. Floudas, *J. Chem. Phys.* **1999**, 110, 9259.
- [15] D. J. Wales, J. P. Doye, M. A. Miller, P. N. Mortenson, T. R. Walsh, *Adv. Chem. Phys.* **2000**, 115, 1.
- [16] K. K. Irikura, R. D. Johnson, *J. Phys. Chem. A* **2000**, 104, 2191.
- [17] E. M. Müller, A. d. Meijere, H. Grubmüller, *J. Chem. Phys.* **2002**, 116, 897.
- [18] M. Dallos, H. Lischka, E. Ventura Do Monte, M. Hirsch, W. Quapp, *J. Comput. Chem.* **2002**, 23, 576.
- [19] J. Baker, K. Wolinski, *J. Comput. Chem.* **2011**, 32, 43.
- [20] P. M. Zimmerman, *J. Comput. Chem.* **2013**, 34, 1385.
- [21] P. M. Zimmerman, *J. Chem. Phys.* **2013**, 138, 184102.
- [22] P. Zimmerman, *J. Chem. Theory Comput.* **2013**, 9, 3043.
- [23] P. M. Zimmerman, *J. Comput. Chem.* **2015**, 36, 601.
- [24] P. M. Zimmerman, *Mol. Simul.* **2015**, 41, 43.
- [25] M. Jafari, P. M. Zimmerman, *J. Comput. Chem.* **2017**, 38, 645.
- [26] A. L. Dewyer, P. M. Zimmerman, *Org. Biomol. Chem.* **2017**, 15, 501.
- [27] A. L. Dewyer, A. J. Argüelles, P. M. Zimmerman, *WIREs Comput. Mol. Sci.* **2017**, e1354.
- [28] D. Rappoport, C. J. Galvin, D. Y. Zubarev, A. Aspuru-Guzik, *J. Chem. Theory Comput.* **2014**, 10, 897.
- [29] B. Schaefer, S. Mohr, M. Amsler, S. Goedecker, *J. Chem. Phys.* **2014**, 140, 214102.
- [30] D. J. Wales, *J. Chem. Phys.* **2015**, 142, 130901.
- [31] S. Habershon, *J. Chem. Phys.* **2015**, 143, 094106.
- [32] S. Habershon, *J. Chem. Theory Comput.* **2016**, 12, 1786.
- [33] M. Bergeler, G. N. Simm, J. Proppe, M. Reiher, *J. Chem. Theory Comput.* **2015**, 11, 5712.
- [34] X.-J. Zhang, Z.-P. Liu, *Phys. Chem. Chem. Phys.* **2015**, 17, 2757.
- [35] L.-P. Wang, R. T. McGibbon, V. S. Pande, T. J. Martinez, *J. Chem. Theory Comput.* **2016**, 12, 638.
- [36] L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande, T. Martínez, *J. Nat. Chem.* **2014**, 6, 1044.
- [37] M. Yang, J. Zou, G. Wang, S. Li, *J. Phys. Chem. A* **2017**, 121, 1351.
- [38] L. D. Jacobson, A. D. Bochevarov, M. A. Watson, T. F. Hughes, D. Rinaldo, S. Ehrlich, T. B. Steinbrecher, S. Vaitheeswaran, D. M. Philipp, M. D. Halls, R. A. Friesner, *J. Chem. Theory Comput.* **2017**, 13, 5780.
- [39] K. Ohno, S. Maeda, *Chem. Phys. Lett.* **2004**, 384, 277.
- [40] S. Maeda, K. Ohno, *J. Phys. Chem. A* **2005**, 109, 5742.
- [41] K. Ohno, S. Maeda, *J. Phys. Chem. A* **2006**, 110, 8933.
- [42] K. Ohno, S. Maeda, *Phys. Scr.* **2008**, 78, 058122.
- [43] S. Maeda, K. Morokuma, *J. Chem. Phys.* **2010**, 132, 241102.
- [44] S. Maeda, K. Morokuma, *J. Chem. Theory Comput.* **2011**, 7, 2335.
- [45] S. Maeda, K. Ohno, K. Morokuma, *Phys. Chem. Chem. Phys.* **2013**, 15, 3683.
- [46] S. Maeda, T. Taketsugu, K. Morokuma, *J. Comput. Chem.* **2014**, 35, 166.
- [47] S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu, K. Morokuma, *Chem. Rec.* **2016**, 16, 2232.
- [48] S. Maeda, Y. Harabuchi, M. Takagi, K. Saita, K. Suzuki, T. Ichino, Y. Sumiya, K. Sugiyama, Y. Ono, *J. Comput. Chem.* **2018**, 39, 233.
- [49] J. A. Varela, S. A. Vazquez, E. Martinez-Nunez, *Chem. Sci.* **2017**, 8, 3843.
- [50] E. Martínez-Núñez, *Phys. Chem. Chem. Phys.* **2015**, 17, 14912.
- [51] E. Martínez-Núñez, *J. Comput. Chem.* **2015**, 36, 222.
- [52] L. J. Broadbelt, S. M. Stark, M. T. Klein, *Ind. Eng. Chem. Res.* **1994**, 33, 790.

- [53] D. M. Matheu, A. M. Dean, J. M. Grenda, W. H. Green, *J. Phys. Chem. A* **2003**, 107, 8552.
- [54] C. W. Gao, J. W. Allen, W. H. Green, R. H. West, *Comput. Phys. Commun.* **2016**, 203, 212.
- [55] P. L. Bhooorasingh, R. H. West, *Phys. Chem. Chem. Phys.* **2015**, 17, 32173.
- [56] P. L. Bhooorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain, R. H. West, *J. Phys. Chem. A* **2017**, 121, 6896.
- [57] Y. V. Suleimanov, W. H. Green, *J. Chem. Theory Comput.* **2015**, 11, 4248.
- [58] X. Ma, W. L. Hase, *Philos. Trans. R. Soc. A* **2017**, 375(2092) DOI: <https://doi.org/10.1098/rsta.2016.0204>.
- [59] D. T. Gillespie, *J. Comput. Phys.* **1976**, 22, 403.
- [60] T. Baer, W. L. Hase, *Unimolecular Reaction Dynamics*; Oxford, **1996**. Oxford University Press.
- [61] MOPAC2016, Stewart, J. J. P., Colorado Springs, CO, USA, **2016**, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net)
- [62] 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, et al., *Gaussian 09*, Wallingford, CT, **2009**.
- [63] O. Tange, *login: The USENIX Magazine* **2011**, 36, 42.
- [64] K. Fukui, *Acc. Chem. Res.* **1981**, 14, 363.
- [65] W. L. Hase, D. G. Buckowski, *Chem. Phys. Lett.* **1980**, 74, 284.
- [66] G. H. Peslherbe, H. Wang, W. L. Hase, *Advances in Chemical Physics*; John Wiley & Sons, Inc., **2007**; pp. 171–201.
- [67] In tsscds2018 wiki: <http://forge.cesga.es/wiki/g/tsscds/HomePage>
- [68] J. Baker, *J. Comput. Chem.* **1986**, 7, 385.
- [69] F. Pietrucci, W. Andreoni, *Phys. Rev. Lett.* **2011**, 107, 085504.
- [70] W. L. Hase, K. Bolton, P. d. Sainte Claire, R. J. Duchovic, X. Hu, A. Komornicki, G. Li, K. F. Lim, D.-H. Lu, G. H. Peslherbe, *Venus05, a General Chemical Dynamics Computer Program*, **2004**.
- [71] E. Rossich Molina, J.-Y. Salpin, R. Spezia, E. Martinez-Nunez, *Phys. Chem. Chem. Phys.* **2016**, 18, 14980.
- [72] E. Martinez-Nunez, D. V. Shalashilin, *J. Chem. Theory Comput.* **2006**, 2, 912.
- [73] S. A. Vazquez, E. Martinez-Nunez, *Phys. Chem. Chem. Phys.* **2015**, 17, 6948.
- [74] R. Perez-Soto, S. A. Vazquez, E. Martinez-Nunez, *Phys. Chem. Chem. Phys.* **2016**, 18, 5019.
- [75] M. J. Wilhelm, E. Martínez-Núñez, J. González-Vázquez, S. A. Vázquez, J. M. Smith, H.-L. Dai, *Astrophys. J.* **2017**, 849, 15.
- [76] J. Zador, A. W. Jasper, J. A. Miller, *Phys. Chem. Chem. Phys.* **2009**, 11, 11040.
- [77] C.-W. Zhou, Z.-R. Li, X.-Y. Li, *J. Phys. Chem. A* **2009**, 113, 2372.
- [78] L. K. Huynh, H. R. Zhang, S. Zhang, E. Eddings, A. Sarofim, M. E. Law, P. R. Westmoreland, T. N. Truong, *J. Phys. Chem. A* **2009**, 113, 3177.
- [79] A. M. El-Nahas, T. Uchimaru, M. Sugie, K. Tokuhashi, A. Sekiya, *J. Mol. Struct. THEOCHEM* **2006**, 770, 59.
- [80] M. Szori, C. Fittschen, I. G. Csizmadia, B. Viskolcz, *J. Chem. Theory Comput.* **2006**, 2, 1575.
- [81] I. Díaz-Acosta, J. R. Alvarez-Idaboy, A. Vivier-Bunge, *Int. J. Chem. Kinet.* **1999**, 31, 29.
- [82] J. R. Alvarez-Idaboy, I. Díaz-Acosta, A. Vivier-Bunge, *J. Comput. Chem.* **1998**, 19, 811.
- [83] D. Ferro-Costas, E. Martínez-Núñez, J. Rodríguez-Otero, E. Cabaleiro-Lago, C. M. Estévez, B. Fernández, A. Fernández-Ramos, S. A. Vázquez, *J. Phys. Chem. A* **2018**, 122, 4790.
- [84] D. G. Truhlar, A. D. Isaacson, G. C. Garret, M. Baer, Eds., CRC: Boca Raton, FL, **1985**; p. 65.
- [85] A. Fernández-Ramos, B. A. Ellingston, B. C. Garrett, D. G. Truhlar, *Reviews in Computational Chemistry*, **2007**; pp. 125–232.
- [86] J. L. Bao, D. G. Truhlar, *Chem. Soc. Rev.* **2017**, 46, 7548.
- [87] T. Yu, J. Zheng, D. G. Truhlar, *Chem. Sci.* **2011**, 2, 2199.
- [88] J. L. Bao, R. Meana-Paneda, D. G. Truhlar, *Chem. Sci.* **2015**, 6, 5866.
- [89] T. Yu, J. Zheng, D. G. Truhlar, *J. Phys. Chem. A* **2012**, 116, 297.
- [90] R. Meana-Pañeda, A. Fernández-Ramos, *J. Chem. Phys.* **2014**, 140, 174303.
- [91] L. E. Rush, P. G. Pringle, J. N. Harvey, *Angew. Chem. Int. Ed.* **2014**, 53, 8672.

Received: 14 February 2018

Revised: 3 May 2018

Accepted: 11 May 2018