Automated exploration of the conformational degrees of freedom along reaction profiles - driving a FASTCAR

Oscar Gayraud[†,‡] and Frédéric Guégan[*,†]

†*Applied Quantum Chemistry Group, E4 Team, IC2MP UMR 7285, Université de Poitiers - CNRS, Poitiers 86073, France*

‡*Spectrometry, Interaction and Theoretical Chemistry team, DCM UMR 5250, Université Grenoble Alpes - CNRS, Grenoble 38058, France*

E-mail: frederic.guegan@univ-poitiers.fr,oscar.gayraud@univ-grenoble-alpes.fr

Abstract

This publication aims at presenting a Python-based workflow designed to enable a fully automatised and exhaustive exploration of the conformational degrees of freedom within the calculation of reaction paths in molecular systems. The proposed strategy focuses on effectively representing the lowest energy conformers for intricate, highly rotatable, and non-intuitive transition states, reagents and products, using existing computational tools. The article presents a workflow that is demonstrated through the application of two chemical reactions: a Diels-Alder reaction involving 4,4-dimethyl-3-methylenepent-1-ene and (E)-dec-5-ene, and an intramolecular acid-catalysed cyclization reaction of polycyclic azocane derivative, which both prove to be challenging to study "by hand". The proposed methodology is expected to be of a great help in the modelling of state-of-the art organic chemistry reactions, whose complexity is ever increasing.

1 Introduction

Theoretical modelling of molecular reactions is a challenging task considering the immense complexity of the experimental situations.[1–4] When computing reaction paths, one is always facing the issue of determining whether the lowest energy profile has been identified or not.[5,6] This is noticeably true regarding the conformational degrees of freedom, whose explicit and non-automatised exploration is only conceivable for small molecular systems. While organic chemistry is a field developing towards complexity, theoreticians often need to simplify the molecular systems they study to be able to unfold reaction paths calculations. Such a mismatch, at a time of steady increase in computational power, is an issue in itself. The course of our interactions with colleagues from experimental chemistry led us to consider it, and to set up a fully automated workflow to allow for an easier, exhaustive exploration of conformational degrees of freedom along reaction profiles. The present article is dedicated to the presentation of this solution, which we label hereafter FASTCAR (Fully Automated Sampling for Transition-state CAlculation and Reaction profiles.)

In a first section, we present the basic tools and philosophy of this Python-based workflow, which heavily relies on the CREST library of Grimme and coworkers,[7] supplemented by an additional library, SpyRMSD,[8] helping in rejecting equivalent structures. The second section is dedicated to an illustration of the functioning of FASTCAR.

We have selected two challenging examples embodying the complexity of molecular systems. The first example involves a Diels-Alder reaction between 4,4-dimethyl-3-methylenepent-1-ene (**1**) and (E)-dec-5-ene (**2**), which has a total of 6,561 potential conformers due to eight rotatable bonds. The second example involves an intramolecular acid-catalysed cyclization reaction of polycyclic azocane derivative, a previously analyzed reaction in our group[9] (Figure 1).

Finally, some development perspective for future versions of FASTCAR are proposed, notice-

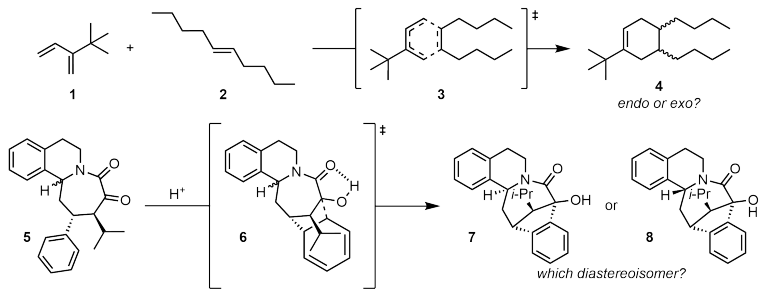

Figure 1: (a) Diels Alder reaction between 4,4-dimethyl-3-methylenepent-1-ene and (E)-dec-5-ene (b) acidic cyclization.

ably related to the elimination of the dependency on the start point geometry that appeared in the case of the intramolecular cyclization reaction.

2 The FASTCAR workflow

As indicated in the introduction, the aim of FASTCAR is to enable the exploration of the conformational degrees of freedom along a given reaction path calculation. It relies mostly on the use of two previously developed tools: CREST[7] and SpyRMSD,[8] interfaced with a DFT code (here Gaussian 16).[10] CREST, developed by the group of S. Grimme, allows the sampling of conformers by the use of DFTB-based metadynamics calculations, either for energy minima (reagents or products) or for transition states (constrained searches). This potent tool proved extremely efficient in producing starting structures for further high-level (DFT) calculations, but in the course of our works we noticed it sometimes struggles to eliminate geometries that should be considered duplicates. Two cases are generally encountered. In the first one, the proposed geometries are actual rotamers of the same conformer with energy and/or rotational constants too high to be considered by CREST as rotamers. This situation is frequently encountered with branched alkyl and phenyl groups; see for instance the case of a tert-butyl rotation on Figure 2.

In the second case, the geometries are artifacts, likely due to inaccuracies in the description of the GFNn-xtb based potential energy surface (for instance two geometries differing by the rotation of a methyl group of an unusual angle) - and which ultimately converge to the

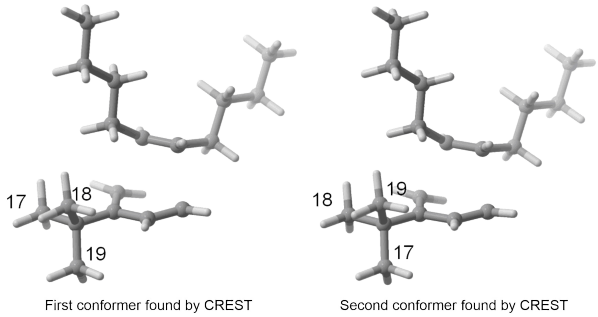

Figure 2: Two conformers found by CREST leading to the same TS after DFT optimisation.

geometry of an already encountered conformer at the DFT optimisation stage. In Figure 3, conformers 11 and 13 both have an invariant RMSD of 0.62 (which exceeds our chosen threshold of 0.5, *vide infra*). However, after DFT optimization, the same transition state is found in both cases.

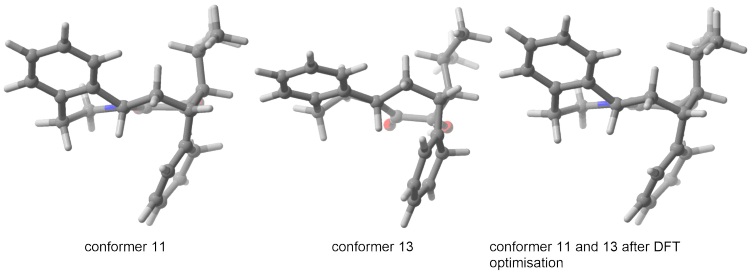

Figure 3: Two conformers leading to the same TS after DFT optimisation

In both cases, the "incorrect" geometries are highly reminiscent of other "correct" conformers found by CREST. This similarity can then be used to set up a refinement in the selection of geometries, which is here performed using SpyRMSD. (Python library for the calculation of invariant Root Mean Square Deviation (RMSD) between structures).

Overall, FASTCAR uses five interrelated Python programs which are detailed hereafter. Note that the current version of FASTCAR supports jobs submission through the Slurm system (version 18.08.8, as used on our local computing cluster).

Automated_CREST:

This first program reads an input file, which contains the parameters to be used for

the CREST search, the SpyRMSD cut-off (see below), the Gaussian calculations to be performed, and the name of a Gaussian output file for the starting geometry. This program then extracts the geometry from the indicated output, checks the nature of calculation to be performed by examining the computed frequencies (constrained TS search if the first frequency is negative, unconstrained otherwise), and starts it using the user-defined parameters and a default energy window of 6 kcal/mol. A first filter for equivalent geometries is applied through the CREGEN utility (using an energy window for rotation barriers of 30 kcal/mol).

Automated_RMSD:

This program further refines the conformer ensemble by applying SpyRMSD, using a user-defined threshold for rejection of rotations-equivalent geometries (default being set to 0.5, in line with former works dealing with conformational analysis[11]). It then initiates Gaussian calculations for each remaining conformer. To save time, frequencies are not calculated in the unconstrained search. If two-step optimisation is chosen, a smaller basis set is used first, e.g. 6-31G(d,p), and after the next part, Automated_Check_1, a larger one, e.g. 6-311++G(d,p), is used.

Automated_Check_1 (optional):

This program filters out unfinished Gaussian calculations, non-optimised geometries and those with incorrect first frequencies, < 0 for geometry and > 0 for TS. In the case of constrained (TS) searches, it further rejects geometries associated to comparable energies and frequencies, which are assumed to be redundant. The criterion for energy is set at 10^{-6} a.u., and at the matching integer value (in cm^{-1}) for frequencies. A final refinement of the TS geometries is achieved by rejection of calculations presenting imaginary frequency less than 0.5 times the one given in input. The resulting conformer ensemble is then used for a subsequent larger-basis set DFT calculation.

Automated_Check_2:

This program continues the workflow by eliminating incomplete and duplicate geometries. In the case of constrained searches it sets a rejection criterion for imaginary modes (less than 0.7 times the negative frequency in the input). If selected, it then initiates an IRC calculation (both forward and backward) followed by geometry optimisation, starting from the geometry with the lowest SCF energy. In the case of unconstrained searches, it starts a frequency calculation and, if selected, additional calculations on the lowest SCF energy geometry.

Automated_EnergyExtract:

This program accomplishes the workflow by extracting the energies of all final calculations. Subsequently, it compiles the extracted energies into a comprehensive text file, facilitating further analysis.

3 Results and discussion

3.1 Computational details and additional note

The electronic structure calculations for all examples presented below were conducted using Gaussian 16 rev B.01.38 software. All calculations were done on a CPU cluster (12-CPU calculations) managed under Slurm 18.08.8, and piloted by FASTCAR. The level of theory used in the following examples is B3LYP-D3/6-31G(d,p) for the Diels-Alder reaction and B3LYP-D3BJ/6-31G(d,p) for the acidic cyclization, chosen as as still representative level in molecular reactivity studies.[12,13] All optimisations conducted under Gaussian were done without constraints. As indicated in the workflow (*vide supra*), vibrational frequencies were systematically computed in the case of the transition state search, and at the end of the search for the minimum case, to ensure in both case the proper nature of the stationary point.

In addition to the examples discussed in the present publication, we additionaly refer the interested reader to an additional work that was submitted concomitantly to this one to

ChemSusChem,[REF] which made extensive use of the FASTCAR approach.

3.2 Diels-Alder reaction with highly rotatable bonds

As a first example, we have selected a Diels-Alder reaction between 4,4-dimethyl-3-methylenepent-1-ene (**1**) and (E)-dec-5-ene (**2**). This reaction was chosen because **2** has 8 rotatable bonds, which generates thousands of possible conformers. Additionally, the presence of a tert-butyl moiety on **1** can reverse the usual endo selectivity. To feed FASTCAR, we have identified an arbitrary transition state (TS) for both the endo and exo geometries without any *a priori*.

For the endo transition state, CREST identified 1258 unique conformers. We reduced this number to 714 by initially pruning with CREGEN, and then applying the SpyRMSD code, which yielded 147 conformers for DFT calculation. Out of these, 144 transition state (TS) calculations were successfully completed within an energy window of 6.2 kcal/mol. The most stable TS was found to be 3.7 kcal/mol lower in energy compared to the arbitrary TS.

For the exo geometry, CREST identified 1562 unique conformers, which were reduced to 909 after an initial pruning step using CREGEN. Finally, 229 conformers were selected for DFT calculation using the SpyRMSD code. A total of 229 transition state (TS) calculations were successfully carried out within an energy window of 7.0 kcal/mol. The comparison between the most stable TS and the arbitrary TS provided revealed that the former was 3.0 kcal/mol lower in energy.

Finally, upon examining the arbitrary TS, it was found that the exo configuration demonstrated greater stability by 1.5 kcal/mol. After the procedure, it was found that the most stable exo conformer had an energy level 0.8 kcal/mol lower than its endo counterpart (Figure 4), showing as expected that a proper consideration of the conformational degrees of freedom is essential if one is desiring to study chemical reactivity and selectivity – used in a two-level Maxwell-Boltzmann model, a difference of activation barriers of 1.5 kcal/mol indeed results in a selectivity at room temperature of 93:7, which drops to 80:20 at 0.8 kcal/mol). On a closing note, the unexpected preference for the exo geometry can be explained by the steric

repulsion caused by the tert-butyl group.

Figure 4: On the left, lowest TS found for endo. On the right, lowest TS found for the exo.

3.3 Intramolecular cyclization of a polycyclic azocane

FASTCAR was then applied to an intramolecular acid-catalysed cyclization reaction of a polycyclic azocane derivative. This reaction had previously been modelled in our group using CREST on the starting materials of both diastereoisomers, to help in reducing the number of starting geometries to consider for the search of transition state (making here use of the Hammond postulate, assuming the lowest energy transition states should relate to the lowest energy conformers for the reagents). This process provided the desired selectivity but at a high cost of human and computational resources. Our aim was to revisit this example using FASTCAR, and check whether a satisfactory reproduction of experimental selectivity was attainable at a much lower cost in terms of human resources.

In this example, CREST identified fewer conformers (19 for **8** and 15 for **7**) due to more constrained geometry. CREGEN did not reduce those ensembles, while SpyRMSD reduced the number of conformers due to duplicates arising mainly from the rotation of the isopropyl group (6 for **8** and **7**). In both cases, we observed a significant decrease in the TS energy compared to the arbitrary TS provided (3.2 kcal/mol for **8** and 9.2 kcal/mol for **7**).

After FASTCAR, it was discovered that the TS for the minor diastereoisomer was 1.1 kcal/mol more stable. However, this finding does not align with the experimental observation (dr over 20:1 in favour of the other). We surmised this discrepancy was related to an incomplete coverage of the potential energy surface by CREST, which "missed" significant regions (FASTCAR still found a TS 3.2 kcal/mol more stable than the input but not the most

stable possible - noticeably the proposed TS was in fact higher in energy than the previously published one). This issue could be addressed by a fine tuning of the search parameters within CREST, to ensure a more thorough exploration of the potential energy surface. However this may be quite complicated to unfold, and the solution likely system-dependent, in conceptual contradiction with the philosophy of FASTCAR. Conversely, by analogy with the approach used in the case of genetic algorithm, we envisioned the possibility to correct this problem by an iterative application of the conformer search.

Two approaches were then attempted to address this discrepancy. The initial attempt was to feed FASTCAR with the most stable conformers found by CREST. This unfortunately resulted in the same conformer ensemble being found in both cases. Subsequently, we focused on the most stable conformer after DFT optimization and selected it as the starting input. After applying FASTCAR for a second time, CREST identified new low energy conformers for the major diastereoisomer. The most stable transition state (TS) was found to be 2.6 kcal/mol lower in energy compared to the TS found during the first iteration. In the case of the minor diastereoisomer, no variation in the conformer ensemble was observed after the reiteration, resulting in the same lowest TS geometry. As a result, a strong selectivity in line with experiment was observed, the difference in activation barriers being in that case of 1.5 kcal/mol (Figure 5). By precaution, FASTCAR was run a third time taking the most stable TS found after the first iteration and to our delight the same lowest TS was found.

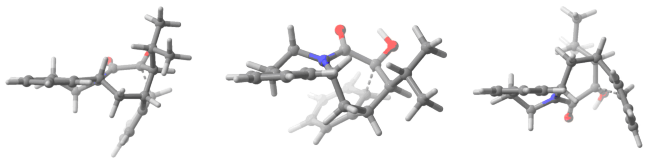

Figure 5: On the left, lowest TS found for major diastereoisomer after the first workflow. On the middle, lowest TS found for major diastereoisomer after a second iteration of the procedure. On the right, lowest TS found for the minor diastereoisomer.

4 Conclusion

In conclusion, FASTCAR is designed to identify the most stable conformers within complex molecular systems using only an arbitrary geometry as a starting point. The workflow streamlines the process of conformer analysis by harnessing the capabilities of established tools, including SLURM, CREST, SpyRMSD, and Gaussian16, making it effortless, rapid, and precise.

The effectiveness of our approach is demonstrated by its successful application to various chemical systems. In the presented examples (along with other works currently under study in our group or being submitted), the transition state located at the end of FASTCAR was more stable than the provided one. In a particular case, FASTCAR needed to be rerun to ensure the location of the global minimum transition state, because of an incomplete exploration of the potential energy surface by CREST. Nevertheless, the presented results demonstrate the ability of the proposed workflow to quickly locate transition states and determine accurate geometries, and forthcoming publications on our group will further illustrate the efficiency of FASTCAR in providing activation and reaction energies meeting experimental data, in various areas from sugar chemistry, organocatalysis and radical chemistry. We believe the solution as it stands is of value for the scientific community, helping in addressing more and more complex chemical reactions by computational methods. In the near future, we will nevertheless improve FASTCAR by the integration of several new features. These will noticeably include the integration of the iterative mode within the main frame, but also the interfacing to other quantum chemistry codes.

References

(1) Cheng, G.-J.; Zhang, X.; Chung, L. W.; Xu, L.; Wu, Y.-D. Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chem-

ical Reactions. *Journal of the American Chemical Society* **2015**, *137*, 1706–1725.

(2) Sperger, T.; Sanhueza, I. A.; Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Accounts of Chemical Research* **2016**, *49*, 1311–1319.

(3) Peng, Q.; Duarte, F.; Paton, R. S. Computing organic stereoselectivity – from concepts to quantitative calculations and predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093–6107.

(4) Reid, J. P.; Sigman, M. S. Comparing quantitative prediction methods for the discovery of small-molecule chiral catalysts. *Nature Reviews Chemistry* **2018**, *2*, 290–305.

(5) Yepes, D.; Neese, F.; List, B.; Bistoni, G. Unveiling the Delicate Balance of Steric and Dispersion Interactions in Organocatalysis Using High-Level Computational Methods. *Journal of the American Chemical Society* **2020**, *142*, 3613–3625.

(6) Burns, M.; Essafi, S.; Bame, J. R.; Bull, S. P.; Webster, M. P.; Balieu, S.; Dale, J. W.; Butts, C. P.; Harvey, J. N.; Aggarwal, V. K. Assembly-line synthesis of organic molecules with tailored shapes. *Nature* **2014**, *513*, 183–188.

(7) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(8) Meli, R.; Biggin, P. C. spyrmsd: symmetry-corrected RMSD calculations in Python. *Journal of Cheminformatics* **2020**, *12*, 49.

(9) Reviriot, Y.; Michelet, B.; Beaud, R.; Martin-Mingot, A.; Guégan, F.; Thibaudeau, S.; Rodriguez, J.; Bonne, D. Hidden Heptacyclic Chiral N-Acyl Iminium Ions: A New Entry to Enantioenriched Polycyclic Azepanes and Azocanes by Superacid-Promoted Intramolecular Pictet-Spengler Reaction. *Chemistry – A European Journal* **2022**, *28*, e202200432.

(10) Frisch, M. J. et al. Gaussian˜16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(11) Ebejer, J.-P.; Morris, G. M.; Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *Journal of Chemical Information and Modeling* **2012**, *52*, 1146–1158.

(12) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(13) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.