Overcoming the Pitfalls of Computing Reaction Selectivity from Ensembles of Transition States

Ruben Laplaza,[†,‡] Matthew D. Wodrich,[†,‡] and Clemence Corminboeuf[*,†,‡]

†*Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland*

‡*National Center for Competence in Research-Catalysis (NCCR-Catalysis), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

Received June 3, 2024; E-mail: clemence.corminboeuf@epfl.ch

Abstract: The prediction of reaction selectivity is a challenging task for computational chemistry, not only because many molecules adopt multiple conformations, but also due to the exponential relationship between effective activation energies and rate constants. To account for molecular flexibility, an increasing number of methods exist that generate conformational ensembles of transition state (TS) structures. Typically, these TS ensembles are Boltzmann weighted and used to compute selectivity assuming Curtin-Hammett conditions. This strategy, however, can lead to erroneous predictions if appropriate filtering of the conformer ensembles is not conducted. Here, we demonstrate how any possible selectivity can be obtained by processing the same sets of TS ensembles for a model reaction. To address the burdensome filtering task in a consistent and automated way, we introduce *marc*, a tool for the modular analysis of representative conformers that aids in avoiding human errors while minimizing the number of reoptimization computations needed to obtain correct reaction selectivity.

Relying on computational methods, such as density functional theory (DFT), to accurately predict reaction selectivity remains a key challenge for *in silico* catalyst design.[1–5] Small errors in computed transition state (TS) energies, even those below chemical accuracy (1 kcal/mol), can result in a reversal of predicted selectivity[6] due to the exponential relationship between effective activation energies and rate constants.[7–11] Dealing with these accuracy issues can further be complicated when large and flexible functional groups used to impart asymmetry through non-covalent interactions are present,[12–14] as these larger systems are likely to adopt multiple TS conformations.

Computational approaches for estimating selectivity often resort to choosing one (or a handful) of relevant conformations derived either from "chemical intuition" or discerned from experimental evidence. The relative free energies of the presumed reaction pathways are then computed and the resulting selectivity estimated.[15–17] While this computationally inexpensive approach may work for simple systems, it becomes increasingly tricky for larger species and cannot be generalized to large pools of catalysts.[18–22] On the other extreme, (*ab initio*) molecular dynamics combined with enhanced sampling techniques such as metadynamics[23] or replica exchange (REMD)[24–27] (which may be powered by machine learning potentials[27–31]) can provide full conformational landscapes that would yield accurate selectivity predictions. In practice, however, such approaches are generally too computationally demanding, either in terms of directly modeling the system over long time frames, or in generating the amount of training data needed to create ML models, and are thus limited to smaller systems.[29,31–33]

One pragmatic approach for determining selectivity from DFT data is to assume a system operates under Curtin-Hammett conditions.[34] In such cases, full conformational sampling of TS structures can be undertaken, and the resulting energies weighted to obtain final product ratios.[16,17,35–40] The popularity of this "Curtin-Hammett Conformational Sampling" (CHCS) method has fostered an increasing number of tools that rely on rotamer libraries,[41–44] inexpensive potentials combined with enhanced sampling techniques[45–47] as popularized by the CREST program[48–52] or distance geometry methods[53–58] to generate conformational ensembles of a molecule. These approaches provide more complete pictures of selectivity, but also require additional computations. As an example of the importance of conformational degrees of freedom, we recently demonstrated how on–the–fly conformational sampling can be used to accurately model enantioselectivity for a diverse set of catalysts with reduced human intervention.[58–60]

When using any of the aforementioned approaches for conformer sampling, particularly the automated variants, inadequate handling of the TS ensembles can lead to significant errors in selectivity estimations. This arises primarily due to two situations: (1) the counting of multiple equivalent transition states (Repeated conformers, Figure 1) and (2) not distinguishing interconvertible and non-interconvertible pathways (Non-interconvertible conformers, Figure 1). Here, we highlight potential pitfalls of using the CHCS strategy by demonstrating how processing the same ensemble of computed TS conformers in various ways leads to virtually any selectivity prediction, even for a simple organic reaction. We then introduce *marc*, a tool for the **m**odular **a**nalysis of **r**epresentative **c**onformers which improves selectivity predictions by untangling conformational ensembles through automated conformer classification and filtering.

Concealed Error Sources in Transition State Conformer Weighting. "Repeated conformer" errors arise when the same (or fundamentally identical) TSs present within an ensemble are counted multiple times. Such errors have different effects depending on how selectivity is determined. In Boltzmann weighting, for instance, a repeated high energy TS can artificially raise the TS barrier height toward that product. On the other hand, if selectivity is assessed directly from rate constants, then a repeated

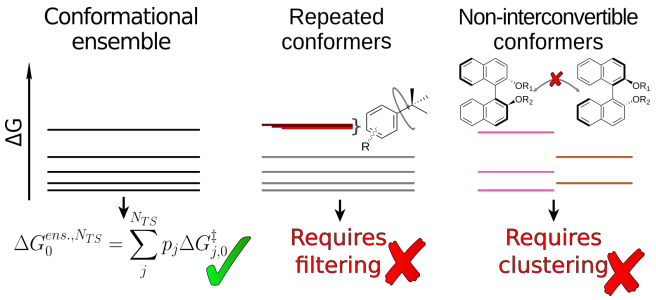


Figure 1. Schematic representation of the two concealed error sources in transition state conformer weighting.

low energy TS can artificially lower the barrier toward that product. In theory, such errors are easily avoided by manually filtering redundant species. Automation, however, introduces its own set of problems as small numerical discrepancies in bond lengths/angles and/or energies, causes symmetry related structures to not be recognized by the program. The equivalence associated with atom indexing (such as that which occurs in rotations of *t*-butyl or phenyl groups) can also lead to the persistence of repeated conformers, even after filtering.

"Interconversion error", the error associated with not distinguishing and properly treating interconvertible and non-interconvertible structures, is subtler and trickier to process. In the potential energy surface, interconvertibility between TSs (1st order saddle points) is governed by temperature-dependent barrier heights (2nd order saddle points), which are hard to characterize.[61] Clearly, two TSs differing only by, for instance, a small rotation of a C−Ph single bond (*i.e.*, rotamers) are connected by a negligibly small energetic barrier, making these species easily interconvertible as they readily adopt the most energetically favorable structure to bypass the TS. On the other hand, conformationally locked structures, such as C_2-symmetric biaryl moieties (Figure 1), cannot intercovert due to a high barrier associated with significant steric repulsion. Conformer generation tools are not necessarily bound by realistic kinetics, which results in the presence of different TS conformers within an ensemble that may not be accessible from one another. In principle, non-interconvertible TSs should be treated as separate reaction pathways, with the rate constants associated with each pathway leading to the same product being summed. On the other hand, interconvertible TSs should be treated as a single pathway. Improper treatment resulting in "double counting" in this setting would lead to an artificial lowering of the effective activation energy.

To illustrate how these error sources quantitatively impact selectivity predictions, we examine the *N*-methylation of a tropane (**1**) with isotopically labeled $^{14}CH_3I$ (Figure 2).[62–65] Two conformations of the system (**1a** and **1b**) exist in equilibrium through a pyramidal inversion of the bridging nitrogen (**TSi**). An S_N2 reaction with $^{14}CH_3I$ leads to two methylated isotopomers (**2a** and **2b**) formed through **TSa** and **TSb**, respectively. As the activation energies associated with the S_N2 reaction are significantly larger ($>$ 12 kcal/mol) than that of the pyramidal N-inversion ($<$ 5 kcal/mol through **TSi**), the system operates under Curtin-Hammett conditions and the product distribution exclusively depends on the free energy barriers of **TSa** and **TSb**. Using this system as an illustrative model,

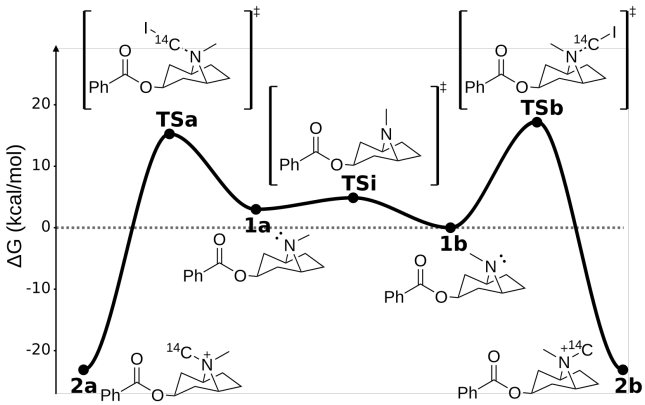


Figure 2. Free energy profile of the *N*-methylation reaction of 3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane with isotopically labeled $^{14}CH_3I$ via S_N2 transition states **TSa** and **TSb**.

we calculate selectivity (expressed as an isotopomer ratio) employing different strategies to account for repeated and (non)interconvetible conformer issues.

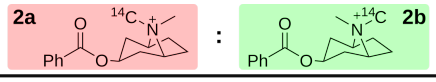
Selectivity Prediction without Conformational Sampling. As a first approximation, we identified structures for **TSa** and **TSb** which were optimized at the ωB97XD/def2-TZVP//ωB97XD/def2-SVP level (see Computational Details for additional information). These computations showed **TSa** to be 2.21 kcal/mol lower in energy than **TSb** (Figure 2). Taking

$$\Delta\Delta G^{\ddagger} = (\Delta G_a^{\ddagger} - \Delta G_b^{\ddagger}) + \Delta G^o = \Delta G_{a,0}^{\ddagger} - \Delta G_{b,0}^{\ddagger}, \quad (1)$$

where the 0 subscript indicates that both activation energies are taken with respect to the lowest energy intermediate (*i.e.* $\Delta\Delta G^{\ddagger}$ is expressed as the difference between the absolute free energies of **TSa** and **TSb**). In this case, $\Delta\Delta G^{\ddagger} = -2.21$ kcal/mol and the major product at $298K$ is predicted to be **2a** following

$$\frac{[\mathbf{2a}]}{[\mathbf{2b}]} = e^{-\Delta\Delta G^{\ddagger}/RT} = 41.547 \approx 98:2 \quad (2)$$

as seen in Figure 3a.



a) Without conformers	98:2
b) Boltzmann weighting (N_{TS}=10)	4:96
c) Boltzmann weighting (N_{TS}=20)	5:95
d) Sum of rate constants (N_{TS}=10)	31:69
e) Sum of rate constants (N_{TS}=20)	49:51
f) Manually curating conformers	4:96
g) Using marc	4:96

Figure 3. Isotopomer ratios calculated with different strategies. a) Without any conformational sampling. b) Through Boltzmann weighting of the 10 lowest TS conformers per pathway. c) As before, but considering the 20 lowest conformers per pathway. d) Through addition of the rate constants from TS conformers, considering the 10 lowest TS conformers per pathway. e) As before, but considering the 20 lowest conformers per pathway. f) Calculated by manually examining and filtering the relevant TS conformers. g) Calculated considering structures selected by *marc*.

Selectivity Prediction with Conformational Sampling. Figure 2 shows that the energies of **TSa** and **TSb** lie close to one another. Structurally, however, both the freely rotating benzoyloxy group and the 8-membered bicycle can adopt a multitude of orientations in the TS. To refine the prediction of selectivity, we performed a constrained conformational search of both **TSa** and **TSb** using CREST version 2.11,[49] (note that the method program used for conformer generation is arbitrary[66]) keeping the relevant I−CH_3−N distances fixed to ensure facile geometric convergence during subsequent ωB97XD reoptimizations. The resulting TS ensembles contained 86 (**TSa**) and 146 (**TSb**) structures (see Figure 4a/b for superimposed structures). We now examine selectivity determined using Boltzmann weighting and summed rate constant approaches.

Boltzmann Weighting of Transition State Conformers. Boltzmann weighting treats TSs as ensembles in which all conformers leading to a specific product are assumed to be freely interconvertible. As a first assumption, we took the $N_{TS} = 10$ lowest energy TS structures leading to each product (as predicted by their GFN2-xTB[67] energies). The TS geometries were reoptimized at the ωB97XD/def2-TZVP//ωB97XD/def2-SVP level (see Computational Details for more information), during which some GFN2-xTB conformers converged to identical TSs (and some similar conformers diverged to different TSs, *vide infra*). We use $\Delta G^{\ddagger}_{j,0}$ to refer to the individual TS energies (relative to **1b**) and add a second subscript to differentiate **a** from **b** when necessary. Boltzmann weighting[68] the ωB97XD TS energies of the 10 lowest energy **TSa** and **TSb** conformers gives the ensemble energies (indicated by the *ens.* superscript which is followed by the number of transition states included, N_{TS}) as

$$\Delta G_0^{ens.,N_{TS}} = \sum_j^{N_{TS}} p_j \Delta G^{\ddagger}_{j,0} \tag{3}$$

where

$$p_j = \frac{e^{-\Delta G^{\ddagger}_{j,0}/RT}}{\sum_j^{N_{TS}} e^{-\Delta G^{\ddagger}_{j,0}/RT}} \tag{4}$$

are the Boltzmann weights. As per eq. 3, $\Delta G_0^{ens.}$ is always higher than the lowest $\Delta G_{j,0}$. The weighting process is conducted separately for all **TSa** and **TSb** conformers.[69] Substituting these values back into eq. 2 yields

$$\Delta\Delta G^{\ddagger}_{ens.,10} = 1.80 \text{ kcal/mol} \tag{5}$$

and

$$\frac{[\mathbf{2a}]}{[\mathbf{2b}]} = \frac{e^{-\Delta G_{a,0}^{ens.,10}/RT}}{e^{-\Delta G_{b,0}^{ens.,10}/RT}} = e^{-\Delta\Delta G^{\ddagger}_{ens.,10}/RT} \approx 4:96 \tag{6}$$

as seen in Figure 3b. Thus, Boltzmann weighting reverses selectivity from the single conformer result (98:2, Figure 3a). Of course, using the lowest 10 energy TS conformers leading to each product is arbitrary, as we do not know the "true" number of unique pathways prior to conformer generation. If we assume 20 conformers to be a better choice and repeat the same process, $\Delta\Delta G^{\ddagger}_{ens.,20} = 1.75$ kcal/mol, giving a 5:95 product ratio (Figure 3d).

Recall that once the lowest energy TS is found, each additional TS conformer identified increases selectivity towards the *opposite* product. As an example, the highest energy **TSb** conformer in our ensemble lies 17.5 kcal/mol above **1b**, which higher than all **TSa** conformers. If duplicates of this high energy conformer are mistakenly added to the ensemble, $\Delta G_{b,0}^{ens.}$ (which is normally lower than $\Delta G_{a,0}^{ens.}$) would slowly tend toward 17.5 kcal/mol, eventually reversing the predicted preferred product. For this reason, duplicate TSs can be problematic, leading to inaccurate selectivity predictions.

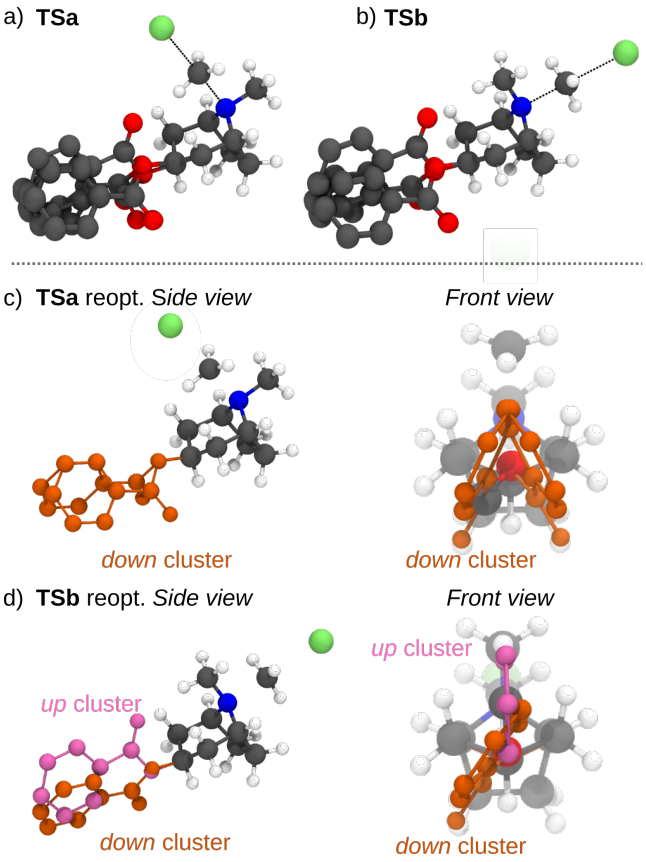


Figure 4. Superimposition of transition state conformers based on the RMSD of the tropane moiety. Full conformational ensemble of a) **TSa** (86 structures), b) **TSb** (146 structures). c) Twenty lowest energy conformers of **TSa** after DFT reoptimization, where all conformers belong to the same (orange) cluster. d) Twenty lowest energy conformers of **TSb** after DFT reoptimization, where conformers belong to two clusters (downward pointing C=O, orange or upward pointing C=O (pink).

Summing Rate Constants of Transition State Conformers. Boltzmann weighting assumes free interconvertability of all TSs leading to a specific product. However, if this interconversion is precluded by a high energetic barrier, then these TSs are best characterized as belonging to different reaction valleys that proceed *in parallel* towards their products. Assuming this is the case for all TS conformers, the effective rate constant $k_{eff.}$ is given as:

$$k_{eff.} \propto e^{-\Delta G^{\ddagger}_{eff.,N_{TS}}/RT} = \sum_j^{N_{TS}} e^{-\Delta G^{\ddagger}_{j,0}/RT} \tag{7}$$

where the mole fraction of the reference state (**1b**) is neglected for simplicity[38] and $-\Delta G_{eff}$ is the effective activation energy towards the product. In other words, the effective kinetic rate constant is now the sum of all individual rate constants for all pathways leading to a specific product.

In this case, the product distribution is the ratio of the total rate constants in each direction, given by

$$\frac{[\mathbf{2a}]}{[\mathbf{2b}]} = \Big[\sum_{j}^{N_{TS}^{a}} e^{-\Delta G_{j,a,0}^{\ddagger}/RT}\Big] / \Big[\sum_{j}^{N_{TS}^{b}} e^{-\Delta G_{j,b,0}^{\ddagger}/RT}\Big] \qquad (8)$$

which is analogous to eqs. 1 and 2 with effective activation energies. Taking $N_{TS}^{a} = N_{TS}^{b} = 10$ and substituting in eq. 8, we obtain $\Delta\Delta G_{eff.,10}^{\ddagger} = 0.48$ kcal/mol and a product ratio [**2a**]/[**2b**] = 0.44, slightly selective towards **2b** (31:69, Figure 3d). Using $N_{TS}^{a} = N_{TS}^{b} = 20$ gives $\Delta\Delta G_{eff.,20}^{\ddagger} = 0.07$ kcal/mol, corresponding to no selectivity (49:51, Figure 3e).

Here, additional low energy conformers significantly accelerate the reaction rate by providing multiple parallel pathways toward the product, while higher energy conformers do not influence the rate. Thus, "double counting" of low energy conformers in this setting lowers the effective activation energy in Equation 7 by adding extra terms. In an extreme case, double counting of the lowest energy TS will give a $RTln(2)$ lower barrier. Importantly, this treatment fundamentally differs from Boltzmann weighting, where high energy conformers would decelerate the reaction.

From the same sets of TS ensembles, we have obtained selectivity predictions ranging from highly selective for **2a** to highly selective for **2b**, solely by post-processing the same results in different ways. To obtain unbiased selectivity predictions, conformer ensemble sorting and selection are required to differentiate interconvertible and parallel TS structures, which obtaining accurate results.

The Right Answer for the Right Reasons. As highlighted above, we lack information about the ability of the different TS conformers in our ensemble to freely interconvert because they are generated without relying on any energetic criterion. For the exemplary case studied here, however, it is possible to manually sort the conformers to arrive at the correct selectivity. Inspecting the DFT-reoptimized **TSa** ensemble reveals only a single conformer family in which all structures are freely interconvertible. This family has a $\Delta G_{a,0,down}^{\ddagger} \approx 15.3$ kcal/mol above **1b** and is characterized most notably by a downward pointing C=O bond (Figure 4c). In contrast, the **TSb** ensemble consists of two distinct conformer families characterized by either upward or downward pointing C=O bonds (Figure 4d), having values of $\Delta G_{b,0,up}^{\ddagger} \approx 17.5$ kcal/mol and $\Delta G_{b,0,down}^{\ddagger} = 13.4$ kcal/mol, respectively.[70] Moving between these two clusters requires overcoming second order saddle points with non-negligible barriers of over 8 kcal/mol above the TSs. As a result, all TSs can be grouped into one of three clusters, either **TSa** with a downward pointing C=O bond, **TSb** with a downward pointing C=O bond, or **TSb** with an upward pointing C=O bond. Using equation 3 we obtain the ensemble energies for each cluster, which can then be separately added, as in Equation 7, to calculate the final selectivity using equation 8. Doing so gives a [**2a**]/[**2b**] ≈ 4:96 (Figure 3f). In the end, only three (of the initial 212) TS conformers actually dictate the selectivity of the reaction.

***marc*: An Automated Clustering Tool to Avoid Errors.** Manually curating structures is time-consuming and unsuitable for all but the simplest systems. To automate this process, we developed *marc* (**m**odular **a**nalysis of **r**epresentative **c**onformers) as a simple command line tool to process conformational ensembles. For a given ensemble, independent of its origin, *marc* uses a mix of geometric (both symmetry-informed heavy-atom RMDSs and dihedral angles) and energetic information (if available) to perform clustering designed to obtain an optimal number of structures needed to completely cover the conformational space. The general workflow of *marc* is shown in Figure 5.

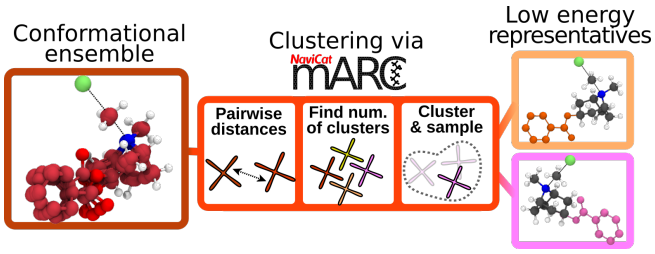


Figure 5. Workflow of *marc*. From a conformer ensemble, *marc* computes pairwise distances using different metrics (heavy atom RMSDs, relative energies, and dihedral angles) to construct a compound distance matrix, finds the optimal number of clusters, and samples the lowest energy structures belonging to each cluster.

Applying *marc* (using the default settings) to the ($N_{TS}^{a} = 20$) DFT-refined ensembles gives a single cluster containing all conformers. The largest heavy-atom RMSD between two structures is a non-negligible 0.87Å, which is sufficiently large to be considered as unique structures based on simple RMSD filtering using predefined thresholds. On the other hand, the maximum energy difference is only 0.01 kcal/mol. By combining energetic and structural criteria, *marc* successfully identifies that all **TSa** conformers as being interconvertible. In contrast, For the ($N_{TS}^{b} = 20$) ensemble is found to have two clusters. The first contains only a single structure with a downwards pointing C=O bond while the second contains the 19 other structures having upward pointing C=O bonds. Here, the maximum heavy-atom RMSD amongst the conformers is slightly larger (1.32Å vs. 0.87Å) while the maximum energy difference is 3.2 kcal/mol with the lowest energy conformer belonging to the first cluster and 19 higher energy conformers belonging to the second. As *marc* selects the same conformers found using manual inspection in the previous section, we once again obtain an isotopomer ratio of 4 : 96 (Figure 3g).

The above results were obtained by processing the 20 lowest energy conformers for both **TSa** and **TSb** optimized at the DFT level. This raises an important question: could the 40 DFT re-optimizations have been completely avoided and the same results obtained? Running *marc* directly on the full CREST-generated ensembles dramatically reduces the number of computations needed. For the original 86 structure **TSa** ensemble, three clusters populated by 1, 1, and 84 structures are identified. The largest corresponds to the downward pointing C=O bond species discussed earlier, while two others contain species with upward pointing C=O bonds that are significantly higher in energy (such that they were not included in the 20 lowest energy structures selected for DFT refinement).[71] For the 146 structure **TSb** ensemble, two clusters characterized by downward (101 structures) and upward pointing C=O bonds (45 structures) are found. If just five total TS structures (one structure from each of the three **TSa** and two **TSb** clusters) are reoptimized using DFT, a 6:94 product ratio is found. This closely matches the 4:96 product ratio obtained by processing the 20 lowest energy DFT reoptimized conformers from **TSa**/**TSb**,

at ∼1/8 of the computational cost.[72] In short, *marc* also simplifies the post-processing of conformational ensembles that have not yet been refined with DFT, reducing large ensembles to a handful of representative structures thereby, reducing computational expense. Such savings become increasingly important when the CHCS protocol in applied to large species and/or in high-throughput settings.

In conclusion, using the *N*-methylation of tropane with isotopically labeled $^{14}CH_3I$ leading to two isotopomers as a model reaction, we have shown how an ensemble of the same TS conformers can be processed in different ways to obtain any possible selectivity prediction under Curtin-Hammett conditions. These different selectivity predictions arise from errors associated with the presence of "repeated conformer" and "interconversion" errors associated with distinguishing when various TSs are freely able to interconvert amongst themselves. We then introduced *marc*, a simple command line tool designed to analyze conformational ensembles and select the most representative structures. Using *marc*, accurate predictions of selectivity can be obtained with significantly reduced computational cost.

Computational Details. Conformer ensemble generation was performed with CREST[48,49,67] version 2.11 using the default settings except doubling the default metadynamics runtime and setting a 0.5 a.u. harmonic constraint placed on the I−CH_3−N atoms involved in the S_N2 transition state. Selected geometries were optimized at the ωB97XD[73]/def2-SVP[74] level of theory as implemented in Gaussian 16.[75] Vibrational frequency analysis was used to confirm that stationary points were either minima (no imaginary frequencies), transition states (one imaginary frequency), or 2nd order saddle points (two imaginary frequencies) on the potential energy surface. Refined energy estimates were obtained by single point computations at the ωB97XD/def2-TZVP level on ωB97XD/def2-SVP geometries. Free energy corrections were taken from the ωB97XD/def2-SVP computations using the GoodVibes program.[37] Solvent effects were included in the single point computations using the SMD[76] implicit solvation model (for acetonitrile). Reported free energies include the solvation-corrected ωB97XD/def2-TZVP electronic energies, and the ωB97XD/def2-SVP free energy corrections. All structures and computed energies are available in the "examples" directory of `https://github.com/lcmd-epfl/marc`.

marc and user instructions can be found at `https://github.com/lcmd-epfl/marc`. Clustering is performed using the kmeans algorithm, with a multidimensional scaling of the averaged dissimilarity matrices as input, as implemented in the scikit-learn python library.[77] The silhouette score method is used to assess the number of clusters. RMSDs are computed considering all isomorphisms between the molecular graphs, which accounts for molecular symmetry.[78] *marc* also allows users to use other clustering algorithms and dissimilarity matrix types, including agglomerative methods and dihedral angles as pioneered by the CENSO program.[68]

Conflicts of Interest. There are no conflicts to declare.

Acknowledgements. The authors are grateful to the EPFL for financial support and computational resources. This publication was created as part of NCCR Catalysis (grant number 180544), a National Centre of Competence in Research funded by the Swiss National Science Foundation. RL thanks Dr. Shubhajit Das and Dr. Simone Gallarati for insightful discussions, and Dr. Marc H. Garner for inspiration regarding acronyms.

Notes and references

(1) Poree, C.; Schoenebeck, F. A Holy Grail in Chemistry: Computational Catalyst Design: Feasible or Fiction? *Acc. Chem. Res.* **2017**, *50*, 605–608.
(2) Ahn, S.; Hong, M.; Sundararajan, M.; Ess, D. H.; Baik, M.-H. Design and Optimization of Catalysts Based on Mechanistic Insights Derived from Quantum Chemical Reaction Modeling. *Chem. Rev.* **2019**, *119*, 6509–6560.
(3) Funes-Ardoiz, I.; Schoenebeck, F. Established and Emerging Computational Tools to Study Homogeneous Catalysis—From Quantum Mechanics to Machine Learning. *Chem* **2020**, *6*, 1904–1913.
(4) Burai Patrascu, M.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N. From desktop to benchtop with automated computational workflows for computer-aided design in asymmetric catalysis. *Nat. Catal.* **2020**, *3*, 574–584.
(5) Fey, N.; Lynam, J. M. Computational mechanistic study in organometallic catalysis: Why prediction is still a challenge. *WIREs Comput. Mol. Sci.* **2021**, e1590.
(6) Krieger, A. M.; Pidko, E. A. The Impact of Computational Uncertainties on the Enantioselectivity Predictions: A Microkinetic Modeling of Ketone Transfer Hydrogenation with a Noyori-type Mn-diamine Catalyst. *ChemCatChem* **2021**, *13*, 3517–3524.
(7) Lam, Y.-h.; Grayson, M. N.; Holland, M. C.; Simon, A.; Houk, K. N. Theory and Modeling of Asymmetric Catalytic Reactions. *Acc. Chem. Res.* **2016**, *49*, 750–762.
(8) Thiel, W. Computational Catalysis-Past, Present, and Future. *Angew. Chem. Int. Ed.* **2014**, *53*, 8605–8613.
(9) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2018**, *119*, 2453–2523.
(10) Harvey, J. N.; Himo, F.; Maseras, F.; Perrin, L. Scope and Challenge of Computational Methods for Studying Mechanism and Reactivity in Homogeneous Catalysis. *ACS Catal.* **2019**, *9*, 6803–6813.
(11) Steiner, M.; Reiher, M. Autonomous Reaction Network Exploration in Homogeneous and Heterogeneous Catalysis. *Top. Catal.* **2022**, *65*, 6–39.
(12) Crawford, J.; Sigman, M. Conformational Dynamics in Asymmetric Catalysis: Is Catalyst Flexibility a Design Element? *Synthesis* **2019**, *51*, 1021–1036.
(13) Cen, S.; Huang, N.; Lian, D.; Shen, A.; Zhao, M.-X.; Zhang, Z. Conformational enantiodiscrimination for asymmetric construction of atropisomers. *Nat. Comm.* **2022**, *13*, 4735.
(14) Maji, R.; Mallojjala, S. C.; Wheeler, S. E. Electrostatic Interactions in Asymmetric Organocatalysis. *Acc. Chem. Res.* **2023**, *56*, 1990–2000.
(15) Wodrich, M. D.; Ye, B.; Gonthier, J. F.; Corminboeuf, C.; Cramer, N. Ligand-Controlled Regiodivergent Pathways of Rhodium(III)-Catalyzed Dihydroisoquinolone Synthesis: Experimental and Computational Studies of Different Cyclopentadienyl Ligands. *Chem. - Eur. J.* **2014**, *20*, 15409–15418.
(16) Peng, Q.; Duarte, F.; Paton, R. S. Computing organic stereoselectivity – from concepts to quantitative calculations and predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093–6107.
(17) Guan, Y.; Wheeler, S. E. Automated Quantum Mechanical Predictions of Enantioselectivity in a Rhodium-Catalyzed Asymmetric Hydrogenation. *Angew. Chem. Int. Ed.* **2017**, *56*, 9101–9105.
(18) Gensch, T.; dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M.; Sigman, M. S.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022**, *144*, 1205–1217.
(19) Alegre-Requena, J. V.; V., S. S. S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S. AQME: Automated quantum mechanical environments for researchers and educators. *WIREs Comp. Mol. Sci.* **2023**, *13*, e1663.
(20) Hashemi, A.; Bougueroua, S.; Gaigeot, M.-P.; Pidko, E. A. HiREX: High-Throughput Reactivity Exploration for Extended Databases of Transition-Metal Catalysts. *J. Chem. Inf. Model.* **2023**, *63*, 6081–6094.
(21) Kammeraad, J. A.; Das, S.; Argüelles, A. J.; Sayyed, F. B.; Zimmerman, P. M. Conformational Sampling over Transition-Metal-Catalyzed Reaction Pathways: Toward Revealing Atroposelectivity. *Org. Lett.* **2024**, DOI:10.1021/acs.orglett.3c04047.
(22) Kalikadien, A. V.; Mirza, A.; Hossaini, A. N.; Sreenithya, A.; Pidko, E. A. Paving the road towards automated homogeneous catalyst design. *ChemPlusChem* **2024**, e202300702.
(23) Fleming, K. L.; Tiwary, P.; Pfaendtner, J. New Approach for Investigating Reaction Dynamics and Rates with Ab Initio

Calculations. *J. Phys. Chem. A* **2016**, *120*, 299–305.
(24) Okamoto, Y.; Sugita, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(25) Petraglia, R.; Nicolaï, A.; Wodrich, M. D.; Ceriotti, M.; C., C. Beyond static structures: Putting forth REMD as a tool to solve problems in computational organic chemistry. *J. Comput. Chem.* **2016**, *37*, 83–92.
(26) Fabregat, R.; A., F.; Meyer, B.; Hollas, D.; C., C. Hamiltonian-Reservoir Replica Exchange and Machine Learning Potentials for Computational Organic Chemistry. *J. Chem. Theory Comput.* **2020**, *16*, 3084–3094.
(27) Gallarati, S.; Fabregat, R.; Juraskova, V.; Inizan, T. J.; Corminboeuf, C. How Robust Is the Reversible Steric Shielding Strategy for Photoswitchable Organocatalysts? *J. Org. Chem.* **2022**, *87*, 8849–8857.
(28) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem. Int. Ed.* **2017**, *56*, 12828–12840.
(29) Young, T. A.; Johnston-Wood, T.; Deringer, V. L.; Duarte, F. A transferable active-learning strategy for reactive molecular force fields. *Chem. Sci.* **2021**, *12*, 10944–10955.
(30) Jurásková, V.; Célerse, F.; Laplaza, R.; Corminboeuf, C. Assessing the persistence of chalcogen bonds in solution with neural network potentials. *J. Chem. Phys* **2022**, *156*, 154112.
(31) Young, T. A.; Johnston-Wood, T.; Zhang, H.; Duarte, F. Reaction dynamics of Diels–Alder reactions from machine learned potentials. *Phys. Chem. Chem. Phys.* **2022**, *24*, 20820–20827.
(32) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today* **2022**, *387*, 143–149.
(33) Margraf, J. T.; Jung, H.; Scheurer, C.; Reuter, K. Exploring catalytic reaction networks with machine learning. *Nat. Catal.* **2023**, *6*, 112–121.
(34) Seeman, J. I. Effect of conformational change on reactivity in organic chemistry. Evaluations, applications, and extensions of Curtin-Hammett Winstein-Holness kinetics. *Chem. Rev.* **1983**, *83*, 83–134.
(35) Suleimanov, Y. V.; Green, W. H. Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.* **2015**, *11*, 4248–4259.
(36) Husch, T.; Seebach, D.; Beck, A. K.; Reiher, M. Rigorous Conformational Analysis of Pyrrolidine Enamines with Relevance to Organocatalysis. *Helv. Chim. Acta* **2017**, *100*, e1700182.
(37) Luchini, G.; Alegre-Requena, J. V.; Funes-Ardoiz, I.; Paton, R. S. GoodVibes: automated thermochemistry for heterogeneous computational chemistry data. *F1000Research* **2020**, *9*, 291.
(38) Williams, I. H. Gibbs energies of activation for reacting systems with multiple reactant-state and transition-state conformations. *J. Phys. Org. Chem.* **2022**, *35*, e4312.
(39) Zhao, Q.; Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nature Comput. Sci.* **2021**, *1*, 479–490.
(40) Zhao, Q.; Hsu, H.-H.; Savoie, B. M. Conformational Sampling for Transition State Searches on a Computational Budget. *J. Chem. Theory Comput.* **2022**, *18*, 3006–3016.
(41) Kothiwale, S.; Mendenhall, J. L.; Meiler, J. BCL::Conf: small molecule conformational sampling using a knowledge based rotamer library. *J. Cheminform.* **2015**, *7*, 47.
(42) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
(43) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O. Rapid virtual screening of enantioselective catalysts using CatVS. *Nat. Catal.* **2018**, *2*, 41–45.
(44) Ingman, V. M.; Schaefer, A. J.; Andreola, L. R.; Wheeler, S. E. QChASM : Quantum chemistry automation and structure manipulation. *WIREs Comput. Mol. Sci.* **2020**, *11*, e1510.
(45) Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O. Prediction of Stereochemistry using Q2MM. *Acc. Chem. Res.* **2016**, *49*, 996–1005.
(46) Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. Application of Q2MM to predictions in stereoselective synthesis. *Chem. Commun.* **2018**, *54*, 8294–8311.
(47) Petraglia, R.; Nicolaï, A.; Wodrich, M. D.; Ceriotti, M.; Corminboeuf, C. Beyond static structures: Putting forth REMD as a tool to solve problems in computational organic chemistry. *J. Comput. Chem.* **2016**, *37*, 83–92.
(48) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
(49) 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.
(50) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551–6568.
(51) Seidel, T.; Permann, C.; Wieder, O.; Kohlbacher, S. M.; Langer, T. High-Quality Conformer Generation with CONFORGE: Algorithm and Performance Assessment. *J. Chem. Inf. Model.* **2023**, *63*, 5549–5570.
(52) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. CREST—A program for the exploration of low-energy molecular chemical space. *J. Chem. Phys.* **2024**, *160*, 114110.
(53) Blaney, J. M.; Dixon, J. S. Distance Geometry in Molecular Modeling. 1994; `https://doi.org/10.1002/9780470125823.ch6`.
(54) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–2474.
(55) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *J. Cheminform.* **2011**, *3*, 8.
(56) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57*, 1747–1756.
(57) Laplaza, R.; Gallarati, S.; Corminboeuf, C. Genetic Optimization of Homogeneous Catalysts. *Chem.–Methods* **2022**, *2*, e202100107.
(58) Sobez, J.-G.; Reiher, M. Molassembler: Molecular Graph Construction, Modification, and Conformer Generation for Inorganic and Organic Molecules. *J. Chem. Inf. Model.* **2020**, *60*, 3884–3900.
(59) Laplaza, R.; Sobez, J.-G.; Wodrich, M. D.; Reiher, M.; Corminboeuf, C. The (not so) simple prediction of enantioselectivity – a pipeline for high-fidelity computations. *Chem. Sci.* **2022**, *13*, 6858–6864.
(60) Wodrich, M. D.; Laplaza, R.; Cramer, N.; Reiher, M.; Corminboeuf, C. Toward in silico Catalyst Optimization. *CHIMIA* **2023**, *77*, 139.
(61) Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angew. Chem. Int. Ed.* **2008**, *47*, 7592–7601.
(62) MacGillavry, C. H.; Fodor, G. 120. The stereochemistry of the tropane alkaloids. Part XVI. The configuration of the N-epimeric ethylmethylnortropinium salts. *J. Chem. Soc.* **1964**, 597.
(63) Bottini, A. T.; la Camp, U. O. D.; Thut, C. C.; Gal, J.; Bellettini, A. G. Stereochemistry of tropane quaternizations. *J. Org. Chem.* **1972**, *37*, 324–326.
(64) Otzenberger, R. D.; Lipkowitz, K. B.; Mundy, B. P. Quaternizations in the 8-azabicyclo[4.3.0]non-3-ene series. *J. Org. Chem.* **1974**, *39*, 319–321.
(65) Babin, V.; Taran, F.; Audisio, D. Late-Stage Carbon-14 Labeling and Isotope Exchange: Emerging Opportunities and Future Challenges. *JACS Au* **2022**, *2*, 1234–1251.
(66) Iribarren, I.; Trujillo, C. Efficiency and Suitability when Exploring the Conformational Space of Phase-Transfer Catalysts. *J. Chem. Inf. Model.* **2022**, *62*, 5568–5580.
(67) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
(68) Grimme, S.; Bohle, F.; Hansen, A.; Pracht, P.; Spicher, S.; Stahn, M. Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules. *J. Phys. Chem. A* **2021**, *125*, 4039–4054.
(69) Note that free energy corrections could be adjusted to account for the number of available microstates (see Reference [37]) which is not consider here.
(70) The fact that we do not find more conformers pointing down resulting from CREST, which uses biased metadynamics, immediately signals that the interconversion between up and down is not trivial.
(71) These structures correspond to the upward pointing conformations of **TSa** that are missing from the energetically pruned ensembles selected for DFT reoptimzation. Reoptimization of these conformers using DFT gave a $\Delta G^{\ddagger}_{a,0,up} \approx 16.2$ kcal/mol for both, well above the 15.3 kcal/mol found for the lowest energy species in the **TSa** ensemble.
(72) The small difference is due to the identification of an additional high-energy **TSa** carbonyl-pointing-up conformer that is missing in the DFT refined ensembles.
(73) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
(74) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
(75) Frisch, M. J. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(76) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
(77) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
(78) Meli, R.; Biggin, P. C. spyrmsd: symmetry-corrected RMSD calculations in Python. *J. Cheminform.* **2020**, *12*, 49.