Emerging conformational-analysis protocols from the RTCONF-16K reaction thermochemistry conformational benchmark set

Bence Balázs Mészáros,[†,‡] Károly Kubicskó,[†,‡] Dávid Dorián Németh,[‡] and
János Daru[*,‡]

[†]*Hevesy György PhD School of Chemistry Eötvös Loránd University,*
Pázmány Péter sétány 1/A, 1117 Budapest, Hungary
[‡]*Department of Organic Chemistry, Eötvös Loránd University,*
Pázmány Péter sétány 1/A, 1117 Budapest, Hungary

E-mail: janos.daru@ttk.elte.hu

Abstract

RTCONF-16K is a new, reactive conformational dataset based on cost-efficient methods to assess different conformational analysis protocols. Our reference calculations underpinned the accuracy of the CENSO (JPCA, 2021, 125, 4039) procedure and resulted in alternative recipes with different cost-accuracy compromises. Our general-purpose and economical protocols (CENSO-light and zero, respectively) were found to be 10-30 times faster than the original algorithm, adding only 0.4-0.7 kcal/mol absolute error to the relative free energy estimates.

1 Introduction

Achieving high-quality results in computational chemistry relies significantly on systematic improbability. The level of electron correlation and basis set can be, in principle, converged systematically following Pople's diagram.[1] Even the accuracy of the popular and cost-efficient Density Functional Theory (DFT) methods can be improved using the concept of Jacob's Ladder[2–4] and extensive benchmark studies,[5–14] albeit the systematic approach is replaced to some extent with empiricism. Furthermore, terms beyond the non-relativistic treatment can be systematically addressed by opening an additional dimension of complexity (Csaszar's cube).[15] Consequently, as of today, accurate electronic structure calculations are possible at the (local) CCSD(T)/CBS level for molecules with over a hundred atoms,[16,17] which can be augmented with cost-efficient thermal contributions via the rigid-rotor and harmonic-oscillator approximation.[18–20] To reach chemical accuracy for free energy estimates,[21] this correction can also be systematically improved by the vibrational perturbation theory.[22]

However, if the thermochemistry of large flexible molecules is considered, free-energy differences between individual configurations cannot explain finite-temperature experiments. Instead, the relative stability of the two *corresponding conformational ensembles* will be relevant.[23–25] With the advent of low-scaling and economical semiempirical methods,[26–29] this principle is gaining popularity and entering the mainstream of static quantum chemistry.[25,30,31]

Although the conformational problem may appear more approachable than electronic-structure calculations, it presents comparable challenges, as pointed out by Grimme et al.[24] Unlike quantum-chemistry methods; however, conformational sampling has no golden standard, results are difficult to systematically improve, its computational cost scales immensely with the increasing number of rotatable bonds,[32] and is often not predictable. Accordingly, many different approaches exist for generating conformers and corresponding energetic ranking, both essential for uncovering the low-energy conformer space. Protocols can be categorized based on their method of conformer generation, which can be done either in a stochastic or systematic manner.[33,34] Stochastic approaches include, for example, molecular dynamics-based,[35–38] Monte Carlo,[39–41] distance geometry,[42–44] and genetic algorithm[45–47] methods. However, these methods generally suffer from a nonpredictable sampling time and the bias of initial geometry.[34] Alternatively, conformers can be generated systematically by varying the torsional angles, but this becomes very soon unfeasible due to the exponential scaling of the problem. More advanced, systematic tools restrain the conformer space by applying cheminformatic rules based on knowledge databases (usually derived from solid-state structure databases) to ensure efficiency.[48–50] However, these rule-based approaches easily fail in the case of unknown systems.[34] Regardless of the method chosen for conformer sampling, employing force-field methods for geometry optimization and energetic ranking of conformers (e.g. AMBER,[51] UFF,[52] MMFF[53]) result in significant errors,[54,55] also they do not apply to transition-metal complexes due to the lack of parametrization.[56] One usual approach is to refine energies obtained from low-level MM/QM calculations on the lowest-energy conformer.[57,58] This can be improved by selecting a larger set of low-energy conformers to refine. Multilevel workflows have been developed for this purpose, leveraging a funnel-like strategy that utilizes a series of methods, each progressively more accurate to predict chemical properties such as $p\mathrm{K_a}$ values.[59–61] One outstanding example of these approaches, the CRENSO protocol enables efficient, general metadynamics-based conformational sampling using CREST,[62,63] and further refinement of the ensemble starting from the semiempirical GFN2-xTB level to more accurate DFT functionals in CENSO.[24,64] This has proven to be a robust tool for the prediction of dynamic molecular properties in solution, such as NMR shift,[65,66] optical rotation,[67] and conformational entropy[68,69] for large flexible molecules.

The accuracy of applied quantum-chemistry methods is a pivotal factor in determining the overall performance of a multilevel protocol. Recent years have seen efforts to assess the performance of many Density Functional Theory (DFT) functionals through benchmark calculations, particularly with regard to relative conformer energies[5,10,24,56] and conformational entropy of individual molecules.[68,70–73] Additionally, some studies have applied conformational analysis to investigate specific chemical reactions, highlighting the significance of including conformational entropy to predict reaction free energies.[68,74,75]

However, despite the above-mentioned advancements, there is no publicly available dataset or systematic study to evaluate the precision of different protocols for predicting reaction free energies, including conformational flexibility. Although different CREST-based methods are emerging for mechanistic investigations of complex organocatalytic reactions,[30,76–80] they face key challenges: they are either expensive or contain simplifications of the protocol, that may introduce unknown and uncontrolled errors, while the computational savings from such simplifications are also unexplored.

Given the instrumental nature of this problem and potentially achievable savings in computational power, we have assembled a versatile reactive conformational thermochemistry benchmark set based on challenging chemical reactions of flexible molecules, where conformational space might significantly change throughout the transformation (e.g., due to ring-closures). These reactions have been supplemented with further organic (ISOL24)[14] and organometallic (MOBH36)[12] reactions to yield our final collection (here termed RTCONF-16K), consisting of 55 diverse chemical reactions with over 16 000 DFT optimized conformers in total (see Supp. for more details and Figure 4 for representative examples). Using this data, we are putting forward new variants of the CENSO protocol that provide large computational savings with moderate loss in accuracy.

2 Methodology

To calculate reaction free energies between two ensembles, we introduce the Gibbs-free energy of a conformational ensemble:[24,68,70,74,81,82]

$$G_{\text{ens}} = \overline{G} - TS'_{\text{conf}}, \tag{1}$$

where $\overline{G}$ denotes the average free energy of individual conformers, T is the thermodynamic temperature, and S'_{conf} is the conformational entropy. This expression can be transformed via a short derivation (see Appendix) to a sum of the free energy of a representative (typically the lowest energy) structure (G^0) and an additive term that describes the entropic stabilization due to the population of multiple conformers ($G^{\mathrm{rel}}_{\mathrm{conf}}$):

$$G_{\mathrm{ens}} = G^0 + G^{\mathrm{rel}}_{\mathrm{conf}}. \tag{2}$$

The latter can be expressed as $G^{rel}_{\mathrm{conf}} = -RT \ln Z^{\mathrm{rel}}$, where Z^{rel} is the partition function defined with the relative free energies of the conformers with respect to the representative configuration. Therefore, to assess the performance of a given conformational free energy estimation protocol, the importance and accuracy of these two terms have to be estimated.

To generate conformational ensembles for each state, we have used the CREST[24] program with GFN2-xTB, in the iMTD-sMTD workflow, applying a 6.0 kcal/mol energy window.[62] The sorting and optimization of the ensemble have been carried out using CENSO[24] together with Orca[83] on different levels of theory (GGA and RSH) according to Jacob's ladder of DFT (see Table 1). For all conformers, thermal contributions have been accounted for on the xTB//xTB level, as it was shown to predict changes in thermostatical correction with good accuracy.[19] Given the fact that solvation of conformational ensembles was found to be crucial for the accurate description of flexible molecules,[69] we applied solvent models for calculations at all levels using DCM as a solvent.

Table 1: Applied electronic structure methods and abbreviations

level of theory	chosen methodology	role
xTB	GFN2-xTB,[26] ALPB(DCM)[84]	conformational search
GGA	B97-3c,[27,85] CPCM(DCM)[86]	higher level SP and OPT
RSH	ωB97M-V[87]/def2-TZVPP,[88] CPCM(DCM)[86]	final SP

As the generation of an exact conformational ensemble is unfeasible, we accept the following protocol as the ground truth for our studies (here termed CENSO-brute-force). In this protocol, all conformers resulting from the CREST calculations have been optimized on the GGA level, and single-point energies were calculated on the RSH level for all conformers. This will be later referred to as the reference (see Table 2). In the method termed CENSO-zero, the lowest-energy conformer according to the xTB-sorted and xTB-optimized ensemble was selected and refined on the RSH//GGA level. In the CENSO-light protocol, the lowest-energy conformer was determined using GGA ranking in the xTB-optimized ensemble.

Table 2: Applied conformational protocols. *Calculated on a narrowed ensemble. **Reference method.

label	ensemble optimization	ensemble ranking	refinement
CENSO-zero	xTB	xTB	RSH//GGA
CENSO-light	xTB	GGA	RSH//GGA
CENSO-default	GGA*	RSH*	RSH//GGA
CENSO-brute-force**	GGA	RSH	RSH//GGA

The original CENSO algorithm (see Figure 1) takes the xTB-optimized ensemble and performs a GGA//xTB level sorting to select conformers within a 4.0 kcal/mol threshold. Next, sequential ensemble optimization is performed, and optimized conformers are filtered based on their GGA//GGA level energies (part2, threshold=2.5 kcal/mol). Finally, the conformer energies within a 99% Boltzmann population sum are refined using RSH singe-point calculations on the part2 ensemble (part3). With the default thresholds applied, we termed this CENSO-default.

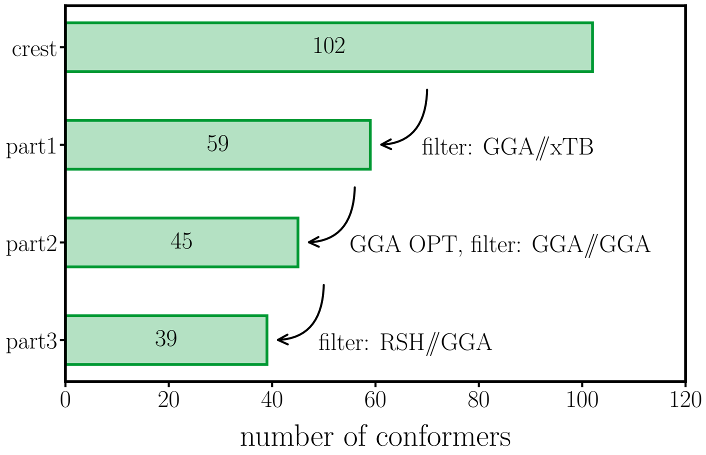

Figure 1: The sequence of different stages in the CENSO-default protocol and the corresponding number of conformers for reaction r34 (reactant and product conformers summed).

3 Results and Discussion

At the beginning of our study, we aimed to stress the importance of incorporating conformational analysis into static reaction modeling. To demonstrate this, we evaluated the errors that may arise from ignoring conformational sampling. We did this by calculating the reaction free energies for the original geometries of the ISOL24 and MOBH35 benchmark sets (which were re-optimized using B97-3c), and then comparing these results to the reference lowest-energy conformers. We also took into account the conformational entropy correction. The resulting errors are quite significant on average (MAE=2.6, RMSE=5.1 kcal/mol), and the maximum error is even more alarming (reaction r39, 20.5 kcal/mol). This is comparable to the additional error of using GGA functionals instead of RSH functionals when compared to local coupled cluster methods.[12] The large errors suggest that chemical-reaction benchmark sets with flexible molecules do not represent the low-energy conformer space. In some cases, the initial geometry bias contributes to unacceptable errors (three reactions that have an absolute error larger than 10 kcal/mol). The corresponding geometrical difference is also illustrated in Figure 2 for the highest error reaction r39, where a ruthenium complex shows significant conformational change after conformational analysis has been performed in both the reactant and the product states (for more details, see SI). This eye-opening finding suggests that comparisons of new methods with these computational benchmark sets implicitly measure performance on conformational and chemical dimensions in an entangled manner.

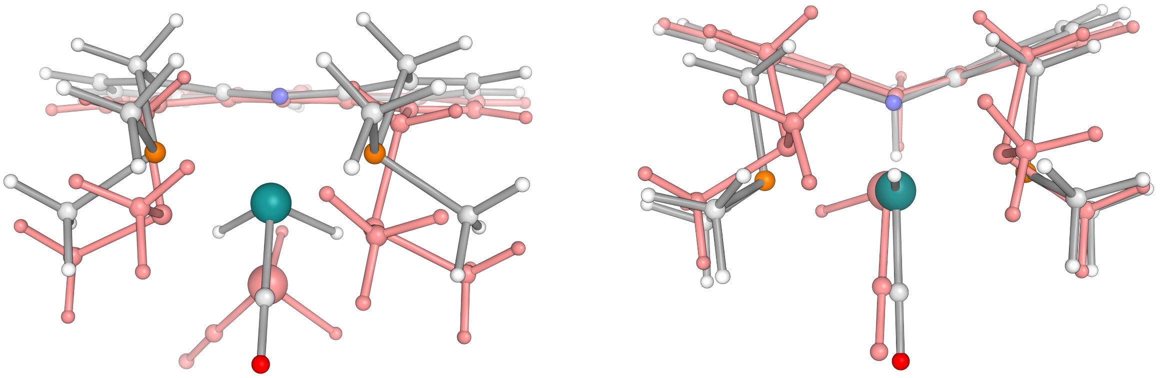
(a) reactant (b) product

Figure 2: The lowest-energy conformers of reactant and product states in r39, overlaid with the original structures (red). Atoms are shown in the respective coloring: teal-ruthenium, orange-phosphorus, red-oxygen, slate-nitrogen, grey-carbon, and white-hydrogen.

Therefore, conformational analysis is essential when describing chemical reactions to accurately describe reaction free energies and conformer geometries. However, the cost of conformational analysis might exceed the cost of demanding local-correlation single-point energy calculations. Therefore, one needs a recipe for efficient management of computational resources. To provide a cost-efficient approach, we have evaluated different protocols for the energetic sorting of the conformers. Additionally, we have also considered refining the lowest-energy conformer for each method.

Compared to the results when conformational analysis was omitted, even rudimentary energetic sorting of the conformers using semiempirical xTB, we observed a significant improvement if the final refinement of the lowest-energy conformer was done on the RSH//GGA level (CENSO-zero, MAE=0.9, MAX=10.5 kcal/mol). Further improvement is possible if the xTB-based conformer ensemble is sorted by GGA single-point energies (CENSO-light, MAE=0.6 kcal/mol, MAX=6.0 kcal/mol). Applying a whole ensemble optimization on the

GGA level, only a small improvement can be achieved compared to the immense cost of the calculation (see Figure 3). Finally, we have assessed the accuracy of a default funnel-like CENSO protocol. It resulted in a MAE of only 0.2 kcal/mol, while reducing the overall cost by half compared to the reference method, allowing impressive cost-efficiency (see Figure 3).

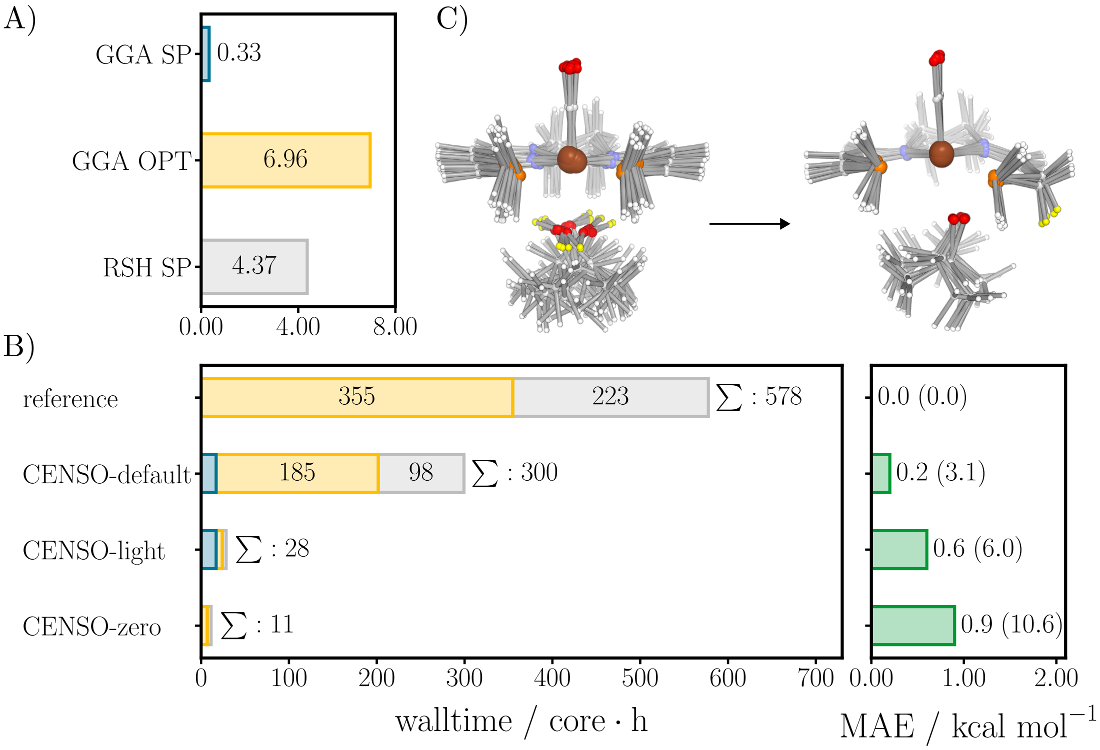

Figure 3: A) Respective timings of a GGA single-point (blue), GGA optimization (yellow), and RSH single-point calculations (gray) per conformer during a CENSO run in reaction r34. B) Timings of different conformer sorting approaches for reaction r34 (102 conformers in total) indicating the various parts of the calculations with the same colors as in panel A. The computational cost of CREST calculations (4.4 core·h) is omitted for clarity. C) 3D structures of the reactant and product states of reaction r34 showing every 4th conformer. Hydrogens that are replaced throughout the reaction are highlighted in yellow. Atoms are shown in the respective coloring: brown-iron, orange-phosphorus, red-oxygen, slate-nitrogen, gray-carbon, and white-hydrogen.

In spite of computational gain through the efficient CENSO algorithm, the relative cost of conformational analysis is still ca. 25 times higher than the optimization and single-point energy calculation of the lowest-energy conformers. Compared to this, the method using GGA//XTB energetic sorting with RSH//GGA refinement on the lowest conformer (CENSO-light) shows a very favorable cost-accuracy balance that provides results with moderate to negligible errors. This method avoids the DFT optimization of the whole (or restricted) ensemble and requires more accurate single-point calculation only for the lowest-energy conformer, which makes it an order of magnitude faster than CENSO-default (see Figure 3) at the cost of 0.4 kcal/mol additional error in MAE (0.6 kcal/mol), and a 2.9 kcal/mol increase in MAX (6.0 kcal/mol). With our present setup, this protocol is only three times more expensive, than omitting conformational analysis. Considering additional error terms in the applied methods (RSH DFT, solvent model, and thermostatistical contributions), CENSO-light can include the conformer ensemble treatment with sufficient accuracy. Alternatively, the CENSO-zero method, which utilizes the xTB//xTB sorting approach, demonstrates a similarly appealing cost-accuracy ratio with lower computational cost. In this case, the computational overhead beyond a standard DFT protocol (geometry optimization and accurate single-point energy evaluation) is minor as it is practically just an additional CREST calculation (4.4 core·h for reaction r34). It is worth mentioning that even this overhead can be sped up by a factor of 5-6 using GPU acceleration.[89] Although this protocol is characterized by larger errors than CENSO-light (MAE=0.9 kcal/mol), its low computational complexity makes it a viable option for quick exploration.

In light of the demonstrated performance of the CENSO-light and zero protocols, the additional computational demand of the CENSO-default can only be justified if final single-point energies are calculated using high-precision (e.g., wave-function) methods to predict reaction free energies. Therefore, in the study of a multistep reaction mechanism, the cost-efficient CENSO-zero/light could be applied to each state to locate rate-determining states,[90] and then, a more demanding CENSO protocol might be applied to refine those states further. This approach offers an additional advantage by circumventing DFT-level ensemble optimization, thus potentially simplifying the conformational search for transition states. Currently, this process is highly laborious due to the need for ensemble TS optimization.[30,76,77] Evaluation of our protocols on TS structures is underway in our laboratory. For reaction mechanism studies, we note that a conformational search protocol is not guaranteed to yield reactive conformers. Therefore, additional effort must be invested if one aims to identify these structures. Similarly, if one is interested in other (e.g., spectroscopic) properties, which depend on the whole ensemble, we recommend the CENSO-default protocol.

Finally, we aimed to examine the role of $\Delta_r G_{\mathrm{conf}}^{\mathrm{rel}}$ term individually in chemical reactions, as there has been no such systematic study to date. Our brute-force reference calculations generally revealed that this term cancels out in most reactions, giving only a 0.8 kcal/mol average absolute value. However, there are significant exceptions, chemically speaking, due to ring transformation or other factors resulting in large changes of conformation degrees of freedom. In particular, in our examples, the ring-closure reactions have shown a positive $\Delta_r G_{\mathrm{conf}}^{\mathrm{rel}}$ between 0-2 kcal/mol (see Figure 4), and somewhat larger values in the double-ring closure of r03 and r06 (1.5 and 1.9 kcal/mol, respectively). Interestingly, for macrocyclic-ring closure of r02 we have obtained a lower $\Delta_r G_{\mathrm{conf}}^{\mathrm{rel}}$ value (0.5 kcal/mol). These results can be understood based on the change in the degree of freedom (DOF) in the rings.[68]

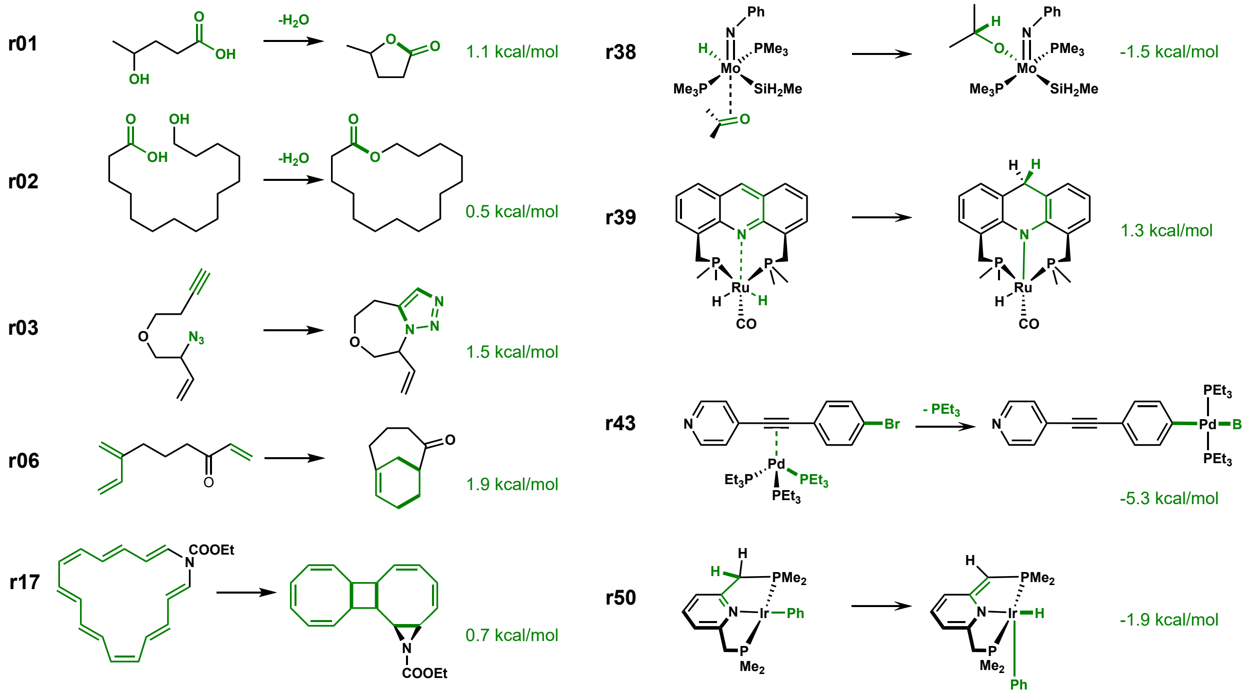

Figure 4: Selected reactions and the corresponding changes of the relative conformational free energies ($\Delta_r G_{\mathrm{conf}}^{\mathrm{rel}}$)

Remarkably, reactions involving the dissociation of a PEt_3 molecule show the largest magnitude of $\Delta_r G^{\mathrm{rel}}_{\mathrm{conf}}$ (r43, -5.3 kcal/mol error and r44, -6.2 kcal/mol error), which can be explained in two terms: First, the large increase in the DOF in PEt_3 after dissociation. Second, the decrease in coordination number and the relief in steric congestion around palladium.

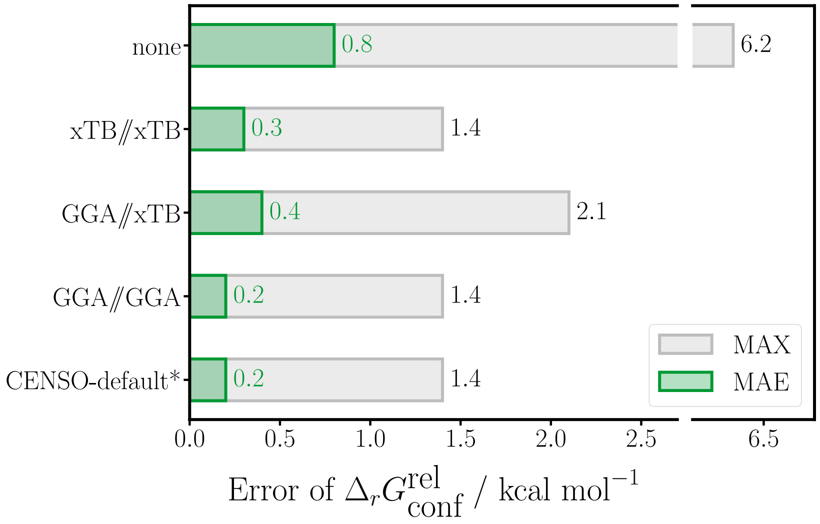

Figure 5: MAE and MAX errors of $\Delta_r G^{\mathrm{rel}}_{\mathrm{conf}}$ using different levels of energetic sorting. *Calculated based on the RSH//GGA level on the reduced ensemble of CENSO

Given that $\Delta_r G^{\mathrm{rel}}_{\mathrm{conf}}$ is a relatively small contribution, one might consider omitting its calculation. Although this approach comes with a moderate 0.8 kcal/mol MAE, the MAX errors are more significant (6.2 kcal/mol). Fortunately, evaluating this term on the cost-efficient xTB//xTB level (as in CENSO-zero), $\Delta_r G^{\mathrm{rel}}_{\mathrm{conf}}$ can be calculated accurately (MAE=0.3 kcal/mol) and importantly, it also shows good performance for outlier reactions, which is indicated by a favorable MAX error (1.4 kcal/mol). With the increase in the applied level of theory, there has been no significant improvement.

4 Conclusions

In a systematic study, we generated a conformational database that features cost-efficient but reasonably-accurate energies for over 16 000 conformers in total for 55 versatile reactions. Based on these well-converged conformational analyses, we have examined the role of different protocols in estimating chemical reaction free energies. Our analysis has shown that the magnitude of $\Delta_r G^{\text{rel}}_{\text{conf}}$ is not significant on average, and a calculation based on xTB already provides an accurate estimation (MAE=0.3 kcal/mol).

Moreover, our investigation highlighted the crucial role of efficiently identifying a representative low-energy conformer in predicting $\Delta_r G_{\text{ens}}$. Additionally, we found that optimizing only the lowest-energy conformer at the GGA level yields a reliable estimate of the geometric relaxation energy for the entire ensemble. As a result, we recommend employing the CENSO-zero protocol, which not only offers a mean absolute error (MAE) of 0.9 kcal/mol but also accelerates computations by 30-40 times compared to the CENSO-default protocol. This recommendation is particularly advantageous when rapid estimations are necessary, especially if GPU acceleration is accessible. The CENSO-light protocol is ideal for general-purpose conformational analysis, offering improved accuracy (MAE=0.6 kcal/mol) while it keeps the computational overhead minimal (10x speed-up compared to CENSO-default). The more sophisticated CENSO-default approach (MAE=0.2 kcal/mol) becomes critical only when the final electronic energies are calculated using high-precision electron-correlation methods or if other properties, e.g., spectroscopic, are required.

Finally, the necessity of conformational analysis for reactions involving flexible molecules has been confirmed. Our study has shown that some flexible-molecule benchmark sets do not represent the low-energy conformer space showing significant errors (MAE=2.6 kcal/mol, MAX=20.5 kcal/mol), comparable to substituting modern range-separated hybrid (RSH) functionals with inferior GGAs. Based on our study, these protocols are a must-have tool for a computational chemist and should be just as routinely applied as RSH functionals.

5 Acknowledgements

We thank Prof. Stefan Grimme for the careful inspection of the manuscript, and his valuable feedback on it. Financial support from the National Research, Development, and Innovation Office (NKFIH, Grant No. FK147031) is gratefully acknowledged. Supported by the ÚNKP-22-2, and ÚNKP-23-3 New National Excellence Programs and the DKOP-23 Doctoral Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

6 Appendix

Here we describe the derivation for the free energy of a conformational ensemble. First, we introduce the partition function $Z = \sum_{j=1}^{n_{\mathrm{rot}}} e^{-\frac{G_j}{RT}}$, and the population of rotamers in the ensemble p_j:

$$p_j = \frac{e^{-\frac{G_j}{RT}}}{\sum_{j=1}^{n_{rot}} e^{-\frac{G_j}{RT}}} \tag{3}$$

Where G_j is the molar free energy of rotamer j. Note that the summation over rotamers might be converted to a conformer-based one, as shown by Pracht and Grimme.[68] Therefore, we decompose the sum over rotamers to sums over subsets of rotamers corresponding to a single conformer

$$\sum_{j=1}^{n_{\mathrm{rot}}} e^{-\frac{G_j}{RT}} = \sum_{j=1}^{g_1} e^{-\frac{G_j}{RT}} + \sum_{j=g_1+1}^{g_1+g_2} e^{-\frac{G_j}{RT}} + ... + \sum_{\sum_1^{i-1} g_i+1}^{n_{\mathrm{rot}}} e^{-\frac{G_j}{RT}}, \tag{4}$$

where the number of rotamers that belong to the conformer i is labeled with g_i. Using that the energies of rotamers corresponding to one conformer are the same due to permutational symmetry, the partition sum simplifies to a sum over n_{conf} conformers:

$$\sum_{j=1}^{n_{\text{rot}}} e^{-\frac{G_j}{RT}} = \sum_{i=1}^{n_{\text{conf}}} g_i e^{-\frac{G_i}{RT}}. \tag{5}$$

Accordingly, the difference between rotamer and conformer populations must be considered. To this end, the conformational entropy can be written according to the Gibbs-Shannon equation. As all states have to be included, this must be done using the populations of individual rotamers (p_j) in a sum over all rotamers:

$$S'_{\text{conf}} = -R \sum_{j=1}^{n_{\text{rot}}} p_j \ln p_j = -R \sum_{j=1}^{n_{\text{rot}}} p_j \ln \left(\frac{e^{-\frac{G_j}{RT}}}{Z} \right) \tag{6}$$

$$= -R \sum_{j=1}^{n_{\text{rot}}} p_j \left(\ln \left(e^{-\frac{G_j}{RT}} \right) - \ln Z \right) \tag{7}$$

$$= R \sum_{j=1}^{n_{\text{rot}}} p_j \frac{G_j}{RT} + R \left(\sum_{j=1}^{n_{\text{rot}}} p_j \right) \ln Z = \frac{1}{T} \sum_{j=1}^{n_{\text{rot}}} p_j G_j + R \ln Z \tag{8}$$

Using the definition of average molar free energy: $\overline{G} = \sum_{j=1}^{n_{\text{rot}}} p_j G_j$.

$$S'_{\text{conf}} = \frac{\overline{G}}{T} + R \ln Z \tag{9}$$

Now substituting into the Gibbs free energy of the ensemble, G_{ens}:

$$G_{\text{ens}} = \overline{G} - T S'_{\text{conf}} = \overline{G} - T \left(\frac{\overline{G}}{T} + R \ln Z \right) = -RT \ln Z \tag{10}$$

From a numerical point of view, we factor out the free energy of a representative conformer G^0 (conveniently, the lowest energy one), and use only relative conformer free energies ΔG_j in the exponent, and name the resulting ΔG_j-based partition function Z^{rel}.

$$Z = \sum_{j=1}^{n_{\text{rot}}} e^{-\frac{G_j}{RT}} = \sum_{j=1}^{n_{\text{rot}}} e^{-\frac{G^0 + \Delta G_j}{RT}} = e^{-\frac{G^0}{RT}} \cdot \left(\sum_{j=1}^{n_{\text{rot}}} e^{-\frac{\Delta G_j}{RT}} \right) = e^{-\frac{G^0}{RT}} \cdot Z^{\text{rel}} \tag{11}$$

Substituting this into G_{ens}, we get

$$G_{\mathrm{ens}} = -RT \ln Z = G^0 - RT \ln Z^{\mathrm{rel}} = G^0 + G^{\mathrm{rel}}_{\mathrm{conf}} \tag{12}$$

Alternatively, one might approximate the partition function using energies instead of free energies, and use thermic correction only on the representative conformer.

$$\sum_{j=1}^{n_{\mathrm{rot}}} e^{-\frac{\Delta G_j}{RT}} \approx \sum_{j=1}^{n_{\mathrm{rot}}} e^{-\frac{\Delta E_j}{RT}} \tag{13}$$

This might save a considerable computational time depending on the applied level of theory. However, it causes a comparable error (MAE=0.4 kcal/mol, MAX=1.5 kcal/mol) to the value of $\Delta_r G^{\mathrm{rel}}_{\mathrm{conf}}$ (average absolute value of 0.8 kcal/mol, See Supporting Information for more details). To address this problem, we added $(G - E)$ corrections on the xTB level for all conformers.[20]

References

(1) Pople, J. Two-dimensional chart of quantum chemistry. *The Journal of Chemical Physics* **1965**, *43*, S229–S230.

(2) Kurth, S.; Perdew, J. P.; Blaha, P. Molecular and solid-state tests of density functional approximations: LSD, GGAs, and meta-GGAs. *International Journal of Quantum Chemistry* **1999**, *75*, 889–909.

(3) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *The Journal of Chemical Physics* **2005**, *123*.

(4) Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; others Accurate first-principles structures and energies

of diversely bonded systems from an efficient density functional. *Nature Chemistry* **2016**, *8*, 831–836.

(5) Rezac, J.; Bím, D.; Gutten, O.; Rulisek, L. Toward accurate conformational energies of smaller peptides and medium-sized macrocycles: MPCONF196 benchmark energy data set. *Journal of Chemical Theory and Computation* **2018**, *14*, 1254–1266.

(6) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive thermochemical benchmark set of realistic closed-shell metal organic reactions. *Journal of Chemical Theory and Computation* **2018**, *14*, 2596–2608.

(7) Neugebauer, H.; Pinski, P.; Grimme, S.; Neese, F.; Bursch, M. Assessment of DLPNO-MP2 Approximations in Double-Hybrid DFT. *Journal of Chemical Theory and Computation* **2023**, *19*, 7695–7703.

(8) Kriz, K.; Novacek, M.; Rezac, J. Non-covalent interactions atlas benchmark data sets 3: Repulsive contacts. *Journal of Chemical Theory and Computation* **2021**, *17*, 1548–1561.

(9) Santra, G.; Martin, J. M. Do double-hybrid functionals benefit from regularization in the PT2 term? Observations from an extensive benchmark. *The Journal of Physical Chemistry Letters* **2022**, *13*, 3499–3506.

(10) Ehlert, S.; Grimme, S.; Hansen, A. Conformational energy benchmark for longer n-alkane chains. *The Journal of Physical Chemistry A* **2022**, *126*, 3521–3535.

(11) Mester, D.; Kállay, M. Double-hybrid density functional theory for core excitations: Theory and benchmark calculations. *Journal of Chemical Theory and Computation* **2023**, *19*, 1310–1321.

(12) Iron, M. A.; Janes, T. Evaluating transition metal barrier heights with the latest

density functional theory exchange–correlation functionals: The MOBH35 benchmark database. *The Journal of Physical Chemistry A* **2019**, *123*, 3761–3781.

(13) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive thermochemical benchmark set of realistic closed-shell metal organic reactions. *Journal of Chemical Theory and Computation* **2018**, *14*, 2596–2608.

(14) Luo, S.; Zhao, Y.; Truhlar, D. G. Validation of electronic structure methods for isomerization reactions of large organic molecules. *Physical Chemistry Chemical Physics* **2011**, *13*, 13683–13689.

(15) Tarczay, G.; Császár, A. G.; Klopper, W.; Quiney, H. M. Anatomy of relativistic energy corrections in light molecular systems. *Molecular Physics* **2001**, *99*, 1769–1794.

(16) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD (T) method: Improved algorithm and benchmark applications. *Journal of Chemical Theory and Computation* **2018**, *14*, 4193–4215.

(17) Liakos, D. G.; Guo, Y.; Neese, F. Comprehensive benchmark results for the domain based local pair natural orbital coupled cluster method (DLPNO-CCSD (T)) for closed- and open-shell systems. *The Journal of Physical Chemistry A* **2019**, *124*, 90–100.

(18) Simón-Carballido, L.; Bao, J. L.; Alves, T. V.; Meana-Pañeda, R.; Truhlar, D. G.; Fernández-Ramos, A. Anharmonicity of coupled torsions: the extended two-dimensional torsion method and its use to assess more approximate methods. *Journal of Chemical Theory and Computation* **2017**, *13*, 3478–3492.

(19) Spicher, S.; Grimme, S. Efficient computation of free energy contributions for association reactions of large molecules. *The Journal of Physical Chemistry Letters* **2020**, *11*, 6606–6611.

(20) Spicher, S.; Grimme, S. Single-point Hessian calculations for improved vibrational frequencies and rigid-rotor-harmonic-oscillator thermodynamics. *Journal of Chemical Theory and Computation* **2021**, *17*, 1701–1714.

(21) Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High accuracy extrapolated ab initio thermochemistry. *The Journal of Chemical Physics* **2004**, *121*, 11599–11613.

(22) Allen, W. D.; Yamaguchi, Y.; Császár, A. G.; Clabo Jr, D. A.; Remington, R. B.; Schaefer III, H. F. A systematic study of molecular vibrational anharmonicity and vibration-rotation interaction by self-consistent-field higher-derivative methods. Linear polyatomic molecules. *Chemical Physics* **1990**, *145*, 427–466.

(23) 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.

(24) 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. *The Journal of Physical Chemistry A* **2021**, *125*, 4039–4054.

(25) Bursch, M.; Mewes, J.-M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angewandte Chemie International Edition* **2022**, *61*, e202205735.

(26) 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. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671.

(27) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *The Journal of Chemical Physics* **2018**, *148*.

(28) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r2SCAN-3c: A “Swiss army knife” composite electronic-structure method. *The Journal of Chemical Physics* **2021**, *154*.

(29) Müller, M.; Hansen, A.; Grimme, S. ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set. *The Journal of Chemical Physics* **2023**, *158*.

(30) Harden, I.; Neese, F.; Bistoni, G. An induced-fit model for asymmetric organocatalytic reactions: a case study of the activation of olefins via chiral Brønsted acid catalysts. *Chemical Science* **2022**, *13*, 8848–8859.

(31) Grimme, S.; Schreiner, P. R. Computational Chemistry: The Fate of Current Methods and Future Challenges. *Angewandte Chemie (International ed. in English)* **2017**, *57*, 4170–4176.

(32) Houk, K.; Liu, F. Holy grails for computational organic chemistry and biochemistry. *Accounts of Chemical Research* **2017**, *50*, 539–543.

(33) Agrafiotis, D. K.; Gibbs, A. C.; Zhu, F.; Izrailev, S.; Martin, E. Conformational sampling of bioactive molecules: a comparative study. *Journal of Chemical Information and Modeling* **2007**, *47*, 1067–1086.

(34) Hawkins, P. C. Conformation generation: the state of the art. *Journal of Chemical Information and Modeling* **2017**, *57*, 1747–1756.

(35) Bruccoleri, R. E.; Karplus, M. Conformational sampling using high-temperature molecular dynamics. *Biopolymers: Original Research on Biomolecules* **1990**, *29*, 1847–1862.

(36) Wu, X.; Wang, S. Self-guided molecular dynamics simulation for efficient conformational search. *The Journal of Physical Chemistry B* **1998**, *102*, 7238–7250.

(37) Taha, H. A.; Castillo, N.; Roy, P.-N.; Lowary, T. L. Conformational studies of methyl

β-D-arabinofuranoside using the AMBER/GLYCAM approach. *Journal of Chemical Theory and Computation* **2009**, *5*, 430–438.

(38) Zimmerman, M. I.; Bowman, G. R. FAST conformational searches by balancing exploration/exploitation trade-offs. *Journal of Chemical Theory and Computation* **2015**, *11*, 5747–5757.

(39) Miteva, M. A.; Guyon, F.; Tuffï¿ 1/2ry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Research* **2010**, *38*, W622–W627.

(40) Wilson, S. R.; Cui, W.; Moskowitz, J. W.; Schmidt, K. E. Applications of simulated annealing to the conformational analysis of flexible molecules. *Journal of Computational Chemistry* **1991**, *12*, 342–349.

(41) Grebner, C.; Becker, J.; Stepanenko, S.; Engels, B. Efficiency of tabu-search-based conformational search algorithms. *Journal of Computational Chemistry* **2011**, *32*, 2245–2253.

(42) Toolkit, D. Daylight chemical information systems. *Inc.: Aliso Viejo, CA* **2007**,

(43) Riniker, S.; Landrum, G. A. Better informed distance geometry: using what we know to improve conformation generation. *Journal of Chemical Information and Modeling* **2015**, *55*, 2562–2574.

(44) Inc., C. C. G. Molecular operating environment (MOE). 2016.

(45) Vainio, M. J.; Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *Journal of Chemical Information and Modeling* **2007**, *47*, 2462–2474.

(46) Li, J.; Ehlers, T.; Sutter, J.; Varma-O'Brien, S.; Kirchmair, J. CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *Journal of Chemical Information and Modeling* **2007**, *47*, 1923–1932.

(47) Liu, X.; Bai, F.; Ouyang, S.; Wang, X.; Li, H.; Jiang, H. Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. *BMC bioinformatics* **2009**, *10*, 1–14.

(48) Hawkins, P. C.; 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. *Journal of Chemical Information and Modeling* **2010**, *50*, 572–584.

(49) Fu, Z.; Li, X.; Merz Jr, K. M. Conformational analysis of free and bound retinoic acid. *Journal of Chemical Theory and Computation* **2012**, *8*, 1436–1448.

(50) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab-Systematic generation of diverse low-energy conformers. *Journal of Cheminformatics* **2011**, *3*, 1–9.

(51) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; others AmberTools. *Journal of Chemical information and Modeling* **2023**, *63*, 6183–6191.

(52) Rappé, A. K.; Casewit, C. J.; Colwell, K.; Goddard III, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.

(53) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry* **1996**, *17*, 490–519.

(54) Kanal, I. Y.; Keith, J. A.; Hutchison, G. R. A sobering assessment of small-molecule force field methods for low energy conformer predictions. *International Journal of Quantum Chemistry* **2018**, *118*, e25512.

(55) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. Comparisons of different force fields in conformational analysis and searching of organic molecules: A review. *Tetrahedron* **2021**, *79*, 131865.

(56) Bursch, M.; Hansen, A.; Pracht, P.; Kohn, J. T.; Grimme, S. Theoretical study on conformational energies of transition metal complexes. *Physical Chemistry Chemical Physics* **2021**, *23*, 287–299.

(57) Hamza, A.; Schubert, G.; Soós, T.; Papai, I. Theoretical studies on the bifunctionality of chiral thiourea-based organocatalysts: competing routes to C- C bond formation. *Journal of the American Chemical Society* **2006**, *128*, 13151–13160.

(58) Probst, N.; Madarász, Á.; Valkonen, A.; Pápai, I.; Rissanen, K.; Neuvonen, A.; Pihko, P. M. Cooperative assistance in bifunctional organocatalysis: Enantioselective mannich reactions with aliphatic and aromatic imines. *Angewandte Chemie International Edition* **2012**, *51*, 8495–8499.

(59) Yesiltepe, Y.; Nuñez, J. R.; Colby, S. M.; Thomas, D. G.; Borkum, M. I.; Reardon, P. N.; Washton, N. M.; Metz, T. O.; Teeguarden, J. G.; Govind, N.; others An automated framework for NMR chemical shift calculations of small organic molecules. *Journal of Cheminformatics* **2018**, *10*, 1–16.

(60) Haworth, N. L.; Wang, Q.; Coote, M. L. Modeling Flexible Molecules in Solution: A p K a Case Study. *The Journal of Physical Chemistry A* **2017**, *121*, 5217–5225.

(61) Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. Multiconformation, density functional theory-based p K a prediction in application to large, flexible organic molecules with diverse functional groups. *Journal of Chemical Theory and Computation* **2016**, *12*, 6001–6019.

(62) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemi-

cal space with fast quantum chemical methods. *Physical Chemistry Chemical Physics* **2020**, *22*, 7169–7192.

(63) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; others CREST—A program for the exploration of low-energy molecular chemical space. *The Journal of Chemical Physics* **2024**, *160*.

(64) Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *Journal of Chemical Theory and Computation* **2019**, *15*, 2847–2862.

(65) Bohle, F.; Grimme, S. Hydrocarbon Macrocycle Conformer Ensembles and 13C-NMR Spectra. *Angewandte Chemie International Edition* **2022**, *61*, e202113905.

(66) Angyal, P.; Hegedüs, K.; Mészáros, B. B.; Daru, J.; Dudás, Á.; Galambos, A. R.; Essmat, N.; Al-Khrasani, M.; Varga, S.; Soós, T. Total Synthesis and Structural Plasticity of Kratom Pseudoindoxyl Metabolites. *Angewandte Chemie International Edition* **2023**, *62*, e202303700.

(67) Bohle, F.; Seibert, J.; Grimme, S. Automated quantum chemistry-based calculation of optical rotation for large flexible molecules. *The Journal of Organic Chemistry* **2021**, *86*, 15522–15531.

(68) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chemical Science* **2021**, *12*, 6551–6568.

(69) Gorges, J.; Grimme, S.; Hansen, A.; Pracht, P. Towards understanding solvation effects on the conformational entropy of non-rigid molecules. *Physical Chemistry Chemical Physics* **2022**, *24*, 12249–12259.

(70) Suarez, E.; Diaz, N.; Suarez, D. Entropy calculations of single molecules by combining the rigid–rotor and harmonic-oscillator approximations with conformational entropy

estimations from molecular dynamics simulations. *Journal of Chemical Theory and Computation* **2011**, *7*, 2638–2653.

(71) Suarez, D.; Díaz, N. Sampling assessment for molecular simulations using conformational entropy calculations. *Journal of Chemical Theory and Computation* **2014**, *10*, 4718–4729.

(72) Díaz, N.; Suárez, D. Toward Reliable and Insightful Entropy Calculations on Flexible Molecules. *Journal of Chemical Theory and Computation* **2022**, *18*, 7166–7178.

(73) Axelrod, S.; Gomez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data* **2022**, *9*, 185.

(74) Chang, C.-e. A.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proceedings of the National Academy of Sciences* **2007**, *104*, 1534–1539.

(75) Verteramo, M. L.; Stenström, O.; Ignjatovic, M. M.; Caldararu, O.; Olsson, M. A.; Manzoni, F.; Leffler, H.; Oksanen, E.; Logan, D. T.; Nilsson, U. J.; others Interplay between conformational entropy and solvation entropy in protein–ligand binding. *Journal of the American Chemical Society* **2019**, *141*, 2012–2026.

(76) 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.

(77) Zhou, H.; Zhou, Y.; Bae, H. Y.; Leutzsch, M.; Li, Y.; De, C. K.; Cheng, G.-J.; List, B. Organocatalytic stereoselective cyanosilylation of small ketones. *Nature* **2022**, *605*, 84–89.

(78) Zhou, H.; Properzi, R.; Leutzsch, M.; Belanzoni, P.; Bistoni, G.; Tsuji, N.; Han, J. T.; Zhu, C.; List, B. Organocatalytic DYKAT of Si-stereogenic silanes. *Journal of the American Chemical Society* **2023**, *145*, 4994–5000.

(79) Grimm, J. A.; Zhou, H.; Properzi, R.; Leutzsch, M.; Bistoni, G.; Nienhaus, J.; List, B. Catalytic asymmetric synthesis of cannabinoids and menthol from neral. *Nature* **2023**, *615*, 634–639.

(80) Cormanich, R. A.; da Silva, G. D. Autobench V1. 0: Benchmarking Automation for Electronic Structure Calculations. *Journal of Chemical Information and Modeling* **2024**,

(81) Zhou, H.-X.; Gilson, M. K. Theory of free energy and entropy in noncovalent binding. *Chemical Reviews* **2009**, *109*, 4092–4107.

(82) Karplus, M.; Ichiye, T.; Pettitt, B. Configurational entropy of native proteins. *Biophysical Journal* **1987**, *52*, 1083–1085.

(83) Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 73–78.

(84) Ehlert, S.; Stahn, M.; Spicher, S.; Grimme, S. Robust and efficient implicit solvation model for fast semiempirical methods. *Journal of Chemical Theory and Computation* **2021**, *17*, 4250–4261.

(85) 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*.

(86) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *The Journal of Physical Chemistry A* **1998**, *102*, 1995–2001.

(87) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of Chemical Physics* **2016**, *144*.

(88) 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. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.

(89) XTB Documentation: Compiling from source. `https://xtb-docs.readthedocs.io/en/latest/development.html#id7`, [Accessed 05-04-2024].

(90) Kozuch, S.; Shaik, S. How to conceptualize catalytic cycles? The energetic span model. *Accounts of Chemical Research* **2011**, *44*, 101–110.