JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Application Note

Ensemble Analyzer: An Open-Source Python Framework for Automated Conformer Ensemble Refinement

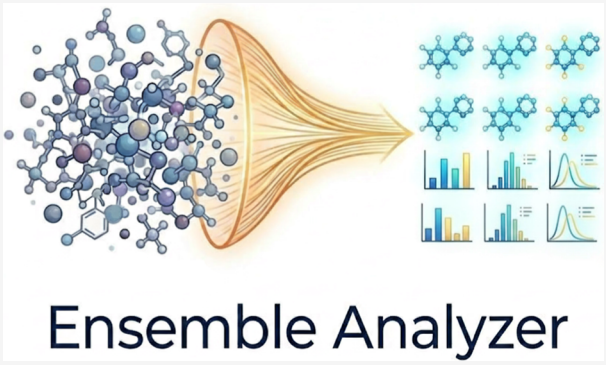
Andrea Pellegrini,* Paolo Righi, Andrea Mazzanti, and Michele Mancinelli

Cite This: J. Chem. Inf. Model. 2026, 66, 5018−5025

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Accurate prediction of molecular properties requires the refinement of the full conformational ensemble at high levels of theory, which remains computationally demanding. Existing workflows typically rely on *ad-hoc* scripts and manual intervention. We present Ensemble Analyzer (EnAn), an open-source Python framework that automates this workflow. EnAn's modular and extensible architecture integrates seamlessly with widely known calculators and allows automated generation and comparison of electronic and vibronic spectra. EnAn effectively manages the reproducible exploration of conformational spaces.



1. INTRODUCTION

Conformational sampling has recently been established as a cornerstone of modern computational organic chemistry.[1−5] The conformational landscape of flexible molecules profoundly influences their physical and chemical properties, and even minimal structure variations might dramatically alter the reactivity, spectroscopic signatures, and molecular assembly. Consequently, the accurate prediction of computed energetics and spectra depends on Boltzmann-weighted contributions of the widest representative ensemble possible. Neglecting this critical step often leads to significant discrepancies between the computed and experimental results.

Over the past five years, the conformational search field has undergone a profound transformation, pushed by the emergence of open-source tools such as RDKit,[6] CREST,[7,8] PyConSolv,[9] and most recently GOAT[10] (integrated within ORCA[11,12]). These software packages have democratized chemical space exploration, enabling a rapid generation of extensive ensembles, often comprising hundreds to thousands of conformers, at minimal computational cost. This efficiency is achieved through relatively inexpensive methods, ranging from classical molecular dynamics to meta-dynamics based on semiempirical extended tight-binding (xTB[13,14]) Hamiltonians. Notably, xTB methods are increasingly becoming among the most widely used for conformational sampling, offering a favorable balance between accuracy and computational speed.

Despite these advances, the significant bottleneck remains in the downstream processing: the refinement of all the conformers at higher levels of theory, required for improving accuracy, remains computationally prohibitive for large ensembles. Conventional protocols have often focused on identifying the single global minimum[15] of the ensemble and carrying it—perhaps few others—over to further higher-level calculations. However, this reductionist approach is inadequate, especially for systems with flat potential energy surfaces, where multiple thermally accessible minima contribute significantly to ensemble-averaged properties and cannot be reliably approximated by a single representative structure.

Bridging the gap between raw ensemble generation and rigorous quantum mechanical (QM) characterization is challenging. Current workflows[16−19] present a rigid infrastructure and/or require the creation of custom scripts to filter structures, manage input/output across different QM engines, and parse results. This technical barrier not only limits the reproducibility of studies but also confines advanced ensemble analysis exclusively to users with strong scripting skills, thus hampering broader adoption of best practices in conformational sampling.

To overcome these limitations, here, we introduce Ensemble Analyzer (EnAn), a flexible and automated software tool designed to streamline the refinement of conformational ensembles. Unlike command-line utilities, EnAn provides an intuitive, user-friendly interface that is accessible to researchers without programming expertise. The software automates the entire workflow: from job submission and output parsing to

Received: February 9, 2026
Revised: April 10, 2026
Accepted: April 13, 2026
Published: April 21, 2026

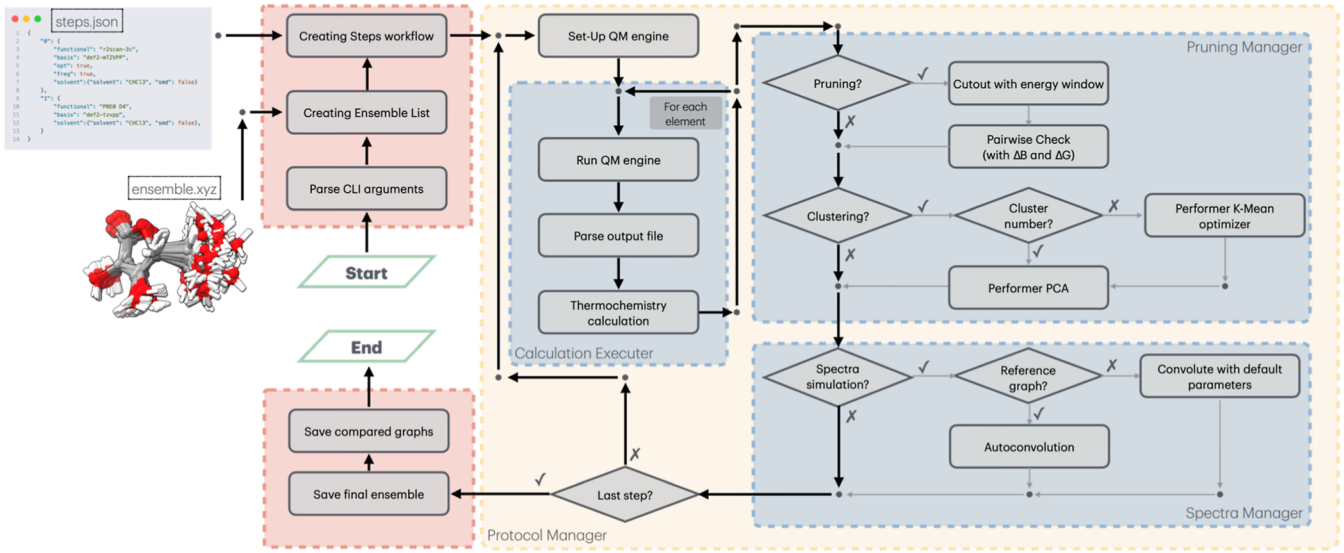


Figure 1. Schematic workflow of Ensemble Analyzer (EnAn).

ensemble pruning, duplicate removal, and dimensionality reduction through Principal Component Analysis (PCA)[20]-guided clustering that preserves structural information and diversity. EnAn further supports the automatic generation, optimization, and comparison of electronic and vibronic spectra, enabling rapid visualization and interpretation of results (see the Supporting Information for a more detailed comparison between EnAn and some of existing tools). By minimizing manual data handling and standardizing workflows, EnAn promotes efficient and reproducible exploration of the complex conformational space.

2. WORKFLOW

EnAn has been developed to automate repetitive quantum chemical calculations over externally generated conformational ensembles. The software architecture follows a modular design based on the Atomic Simulation Environment[21] framework, enabling seamless integration of multiple quantum chemistry engines. The current version implementation supports the implementation of ORCA (version 5.0 and above) and Gaussian16,[22] with the modular structure allowing future extensions to additional QM engines.

The complete workflow (illustrated schematically in Figure 1) requires only two input files: (i) the ensemble geometries in standard XYZ format and (ii) a protocol file defining the computational steps. The protocol is formatted as a JSON file that specifies the sequential calculation steps, each with customizable parameters including level of theory, solvent model, optimization and frequency switches, and pruning thresholds. To streamline protocol creation, an interactive script is provided, guiding users through the setup with three complexity levels: Basic, Intermediate, and Advanced.

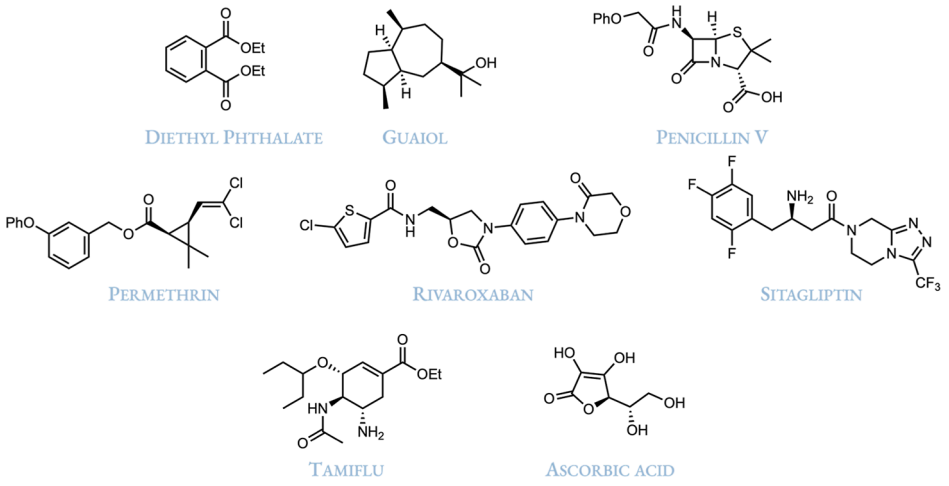
EnAn implements Boltzmann weighting of conformers based on Gibbs free energies evaluated using the quasi-Rigid Rotor Harmonic Oscillator (qRRHO) approximation with Grimme's damping scheme for low-frequency modes.[23] To ensure consistency across different QM engines, EnAn internally recalculates partition functions and thermodynamic contributions by extracting raw data needed for these calculations from output files and applying Grimme's qRRHO formulas uniformly, regardless of the backend computational engine. Relative populations are automatically computed at the specified temperature and used to weight individual conformer contributions to ensemble-averaged properties and spectra.

To efficiently manage the ensemble size and remove redundant structures, EnAn implements a computationally lightweight dual-filter pruning strategy. Unlike traditional methods that rely primarily on Cartesian Root-Mean-Square Deviation (RMSD) alignments, which require expensive geometry alignments of structures, EnAn uses the scalar norm of the rotational constant vector (B) and the energy (Gibbs energy G, when available; otherwise, electronic energy E is used) as the primary filtering metrics. Two conformers are considered identical—and the one with higher relative energy discarding—if they simultaneously satisfy both an energetic ($|E_i - E_j| < thrG$) and a geometric ($|B_i - B_j| < thrB$) equivalence condition. This approach conceptually parallels the fast-filtering heuristics employed by conformer generators, such as CREST, but it is strictly implemented as a pairwise scalar comparison to bypass structural alignments entirely during the refinement steps. The pruning thresholds (thrG, thrB, and $thrG_{max}$) are dynamically selected based on the nature of the computational step. A detailed description of the filtering algorithm, along with the standard default parameters used for all job types, is provided in Section 1.1 of the Supporting Information.

Optional PCA-based clustering can be activated on demand through the protocol settings. This PCA analysis operates on the eigenvalues of atomic distance matrices computed from Cartesian coordinates with an option to include or exclude hydrogen atoms. This flexibility allows users to focus on heavy-atom skeletal differences or include all atoms for a complete conformational description. When clustering is enabled, conformers are grouped using K-means clustering, with either automatic determination of optimal cluster number via silhouette score maximization or user-defined cluster count. For each cluster, the representative structure is automatically selected as the lowest-energy conformer assigned to that cluster centroid. Initial cluster centroids are randomly seeded and then iteratively optimized through the K-means algorithm to maximize the silhouette score, ensuring both energetic

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025

Table 1. Refinement Process of Highly Flexible Molecules[a]



	starting ensemble	protocol 1	protocol 2	protocol 3	final retention rate
Diethylphthalate	156	151	35	35	22%
Guaiol	118	109	34	34	29%
Penicillin V	136	71	12	12	9%
Permethrin	304	304	101	101	33%
Rivaroxaban	110	110	30	30	27%
Sitagliptin	428	343	142	142	33%
Tamiflu	836	527	100	99	12%
Ascorbic acid	134	68	45	45	33%

[a]The number represents the number of conformers active after each protocol. Protocol 1: single point at the g-xTB level, Protocol 2: optimization and frequency calculation at r^{2}SCAN-3c, and Protocol 3: single-point calculation at ωB97X-D4rev/def2-QZVPP.

favorability and conformational diversity in the reduced ensemble.

For spectroscopic applications, EnAn automatically generates ensemble-averaged IR, vibrational circular dichroism (VCD), UV, and electronic circular dichroism (ECD) spectra from the Boltzmann-weighted conformer ensemble when frequency calculations or electronic state energies are requested. Individual conformer spectra are first convoluted using appropriate line shape (LS) functions (Lorentzian for IR/VCD and Gaussian for UV/ECD) with theory-dependent broadening parameters, then combined according to their population weights. When experimental reference spectra are provided (as simple two-column XY data files), EnAn implements an automated optimization routine to determine the optimal spectral alignment parameters. The algorithm simultaneously optimizes both spectral shift (multiplicative scaling for IR/VCD wavenumbers or additive shift for UV/ECD energies) and Full-Width at Half-Maximum (FWHM) of the convolution function by minimizing the RMSD between calculated and experimental spectra within the recorded experimental window. This automated fitting procedure, performed using the L-BFGS-B[24] optimization method with user-definable boundaries, eliminates manual spectral manipulation and provides quantitative similarity metrics between theory and experiment. Optimized spectra, along with fitting parameters and similarity scores, are saved for each protocol step, enabling a systematic evaluation of how the level of theory affects spectral predictions.

3. USAGE

3.1. Ensemble Refinement

Our first case study demonstrates a multilevel automated refinement workflow, illustrating the typical process for ensemble refinement and the subsequent reduction of conformers. Using CREST at the GFN2-xTB level, we have generated the ensembles for several medium to large, highly flexible molecules: Diethyl phthalate, Guaiol, Penicillin V, Permethrin, Rivaroxaban, Sitagliptin, Tamiflu, and Ascorbic acid. Each ensemble was then subjected to a three-step refinement protocol: (i) single point at the g-xTB[25] level, (ii) geometry optimization at the r^{2}SCAN-3c[26] level, and (iii) a single-point electronic energy computation at the ωB97X-D4rev[27,28]/def2-QZVPP[29] level. All calculations were performed using ORCA 6.1.0. Default thresholds were applied, consistent with CREST parameters for identifying new conformers, and the maximum energy window was chosen (3.50 kcal/mol after optimization and frequency calculation) to include all conformers with significant Boltzmann populations. The initial single-point energy window was expanded to account for high-energy structures that might undergo significant stabilization during following geometry optimizations. Specifically, single-point calculations employ a wider energy threshold (ΔG_{max} = 6 kcal/mol) compared to optimization steps, allowing potentially relevant conformers to be carried forward before refinement. All threshold parameters can be manually adjusted in the protocol definition file by the user. Detailed default energy windows for each job type are provided in the Supporting Information (Table S2). The results of the refinement process are summarized in Table 1.

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025

3.2. Principal Component Analysis (PCA) Clustering

As illustrated in the previous example, conformational ensembles can become exceedingly large, making even single-point calculations on all conformers computationally demanding. To mitigate this computational cost, geometry clustering provides an effective strategy to reduce the number of conformers while preserving an effective description of the overall conformational space.

To evaluate the impact of clustering discretization on ensemble-derived properties, three flexible molecules were selected as representative case studies: Vancomycin, Paclitaxel, and Prostaglandin $PGF_{2\alpha}$ (see the Supporting Information Figure S2). Initial ensembles were generated using CREST, and the conformer energies were subsequently re-evaluated at the g-xTB level directly within EnAn. Each ensemble was then clustered while the number of resulting clusters. For each new clustered ensemble, the Boltzmann-weighted average energy was calculated and compared with that obtained from the full ensemble. The resulting trend, shown in Figure 2, reports the

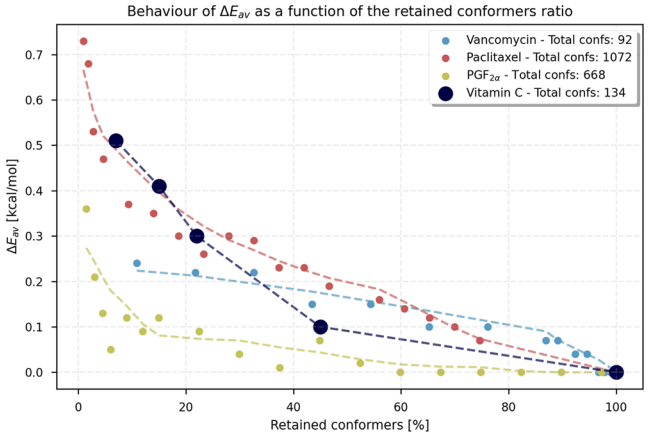


Figure 2. Variation of ΔE_{av} loss as a function of the conformer's retention.

variation of the energy deviation (ΔE_{av}, as defined in eqs 1−3) as a function of the percentage of conformers retained—thus illustrating the energy loss introduced by the progressive reduction in the number of clusters.

$$E_{av} = \sum_{i}^{N} p_i E_i \tag{1}$$

$$p_i = \frac{g_i \exp\left\{-\frac{\Delta E_i}{RT}\right\}}{\sum_j g_j \exp\left\{-\frac{\Delta E_j}{RT}\right\}} \tag{2}$$

$$\Delta E_{av} = E_{av,TOT} - E_{av,clustered} \tag{3}$$

Ascorbic acid was also included; in this case, the values plotted were computed using Gibbs free energy, with electronic energy evaluated at the ωB97X-D4rev/def2-QZVPP level and the thermal correction calculated at the r^2SCAN-3c level.

As expected, reducing the number of conformers decreases the accuracy, as reflected by the increase in ΔE_{av}. Conversely, fewer conformers significantly reduce the wall time required for optimization and frequency calculation (Supporting Information Figure S4 shows how the number of clusters affects the accuracy and wall time of the computation). Therefore, clustering the ensemble represents an excellent strategy for reducing the dimensionality of the conformational space and, consequently, the computational cost. However, it must be applied judiciously to balance efficiency and accuracy. To identify the optimal clustering granularity, EnAn employs an automated optimization strategy. The algorithm systematically evaluates all integer values of n (number of clusters) between 10% and 80% of the ensemble size, with a lower bound of two clusters. For each n value, K-means clustering is performed, and the corresponding silhouette score is calculated. The configuration yielding the highest silhouette score is selected as the optimal clustering solution.

3.3. Spectra Simulation—Electronic and Vibrational Spectra

EnAn has been designed to manage also the most time-consuming aspects of multiconformer spectra simulation. EnAn can generate the electronic and vibrational spectra for all active conformers, weighting each contribution according to the Boltzmann population. Each output obtained from the QM engine is scaled by the conformer's population and then convoluted with the appropriate LS for the corresponding spectrum. Users can customize the default convolution parameters for both LSs separately via the CLI flag "fwhm-vibro" and "fwhm-electro", respectively. In the code, these LSs are implemented as shown in eqs 4 and 5

$$L(\omega_i, \gamma) = \sum_i I_i \frac{\gamma_i^2}{\gamma^2 + 4(X - \omega_i\Delta)^2} \tag{4}$$

$$G(\omega_i, \sigma) = \sum_i I_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X - (\omega_i + \Delta)}{\sigma}\right)^2\right] \tag{5}$$

where $L(\omega_i\gamma)$ denotes the Lorentzian convolution function, I_i is the intensity of the impulse, γ is the FWHM, and X is the entire frequency range considered for the convolution; $G(\omega_i\sigma)$ denotes the Gaussian convolution function, where σ is the broadening parameter calculated as $\sigma = \frac{\gamma}{(2\sqrt{2\ln(2)})}$. In both equations, Δ is the spectral shift applied to align the computed spectrum with the experimentally recorded spectrum.

If experimental spectra are present in the same folder, then the autoconvolute function is triggered. This routine optimizes the convolution parameters to achieve the best overlap between computed and experimental spectra using a bounded numerical minimization (via the minimize function from the scipy.optimize module[30]). The objective is to reduce the RMSD (eq 6) between the two spectra, yielding a similarity index[31,32] as defined in eq 7, where max represents the maximum value of RMSD possible (1 for the achiral spectrum and 2 for the chiral ones). The weighting function $w(x)$ focuses the action of the optimizer and is defined in eq 8.

$$\mathrm{RMSD} = \frac{\sum_i (y_i - y_{exp})^2 w(x_i)}{\sum_i w(x_i)} \tag{6}$$

$$S = 1 - \frac{\mathrm{RMSD}}{\max} \tag{7}$$

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025

$$w(x) = \begin{cases} 0 \text{ if } x < x_{\text{exp}}^{\text{min}} \text{ or } x > x_{\text{exp}}^{\text{max}} \\ \max\left(0.15, \dfrac{g(x_{\text{int}}^{\text{min}}, \sigma_1)}{\max(g(\min x_{\text{int}}, \sigma_1))}\right) \text{ if } x_{\text{exp}}^{\text{min}} < x < x_{\text{int}}^{\text{min}} \\ \max\left(0.15, \dfrac{g(x_{\text{int}}^{\text{max}}, \sigma_2)}{\max(g(\max x_{\text{int}}, \sigma_2))}\right) \text{ if } x_{\text{int}}^{\text{max}} < x < x_{\text{exp}}^{\text{max}} \\ 1 \text{ if } x_{\text{int}}^{\text{min}} \le x \le x_{\text{int}}^{\text{max}} \end{cases} \tag{8}$$

$$g(x_0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right) \tag{9}$$

$$\sigma_1 = 0.1|x_{\text{exp}}^{\text{min}} - x_{\text{int}}^{\text{min}}| \text{ and } \sigma_2 = 0.1|x_{\text{exp}}^{\text{max}} - x_{\text{int}}^{\text{max}}| \tag{10}$$

where $x_{\text{exp}}^{\text{min}}$ and $x_{\text{exp}}^{\text{max}}$ define the lower and upper boundaries of the experimental window and $x_{\text{int}}^{\text{min}}$ and $x_{\text{int}}^{\text{max}}$ define the borders of the region of interest. This region can be customized by the user; if switched off, the weighting function defaults to a step function, assuming a value of 0 outside the experimental window and 1 inside it.

To validate these features, we have resimulated two previously published, feature-rich ECD spectrum of (*S*)-3-(4-bromophenyl)-3-(trifluoromethyl)oxirane-2,2-dicarboxylate[33] and the VCD spectrum of (*S*)-(1,1,1-trifluoro-3-nitropropan-2-yl)benzene.[34]

The ECD of the epoxide has been calculated in acetonitrile using SMD[35] implicit solvation theory, using CAM-B3LYP[36]-D4[37] and ωB97X-D4rev as functionals and def2-TZVPP as the common basis set. The resulting spectrum is shown in Figure 3.

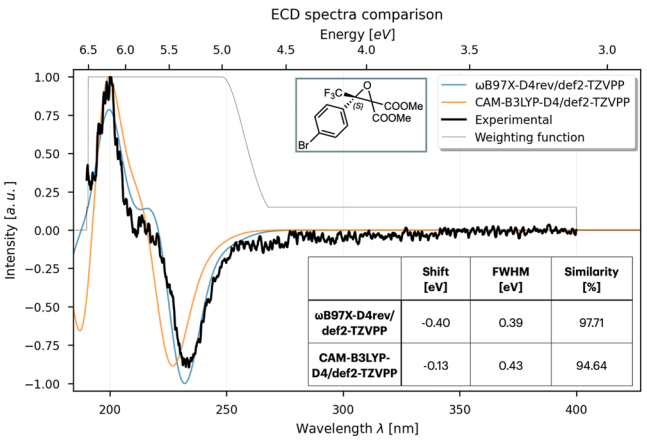


Figure 3. ECD spectrum of (*S*)-3-(4-bromophenyl)-3-(trifluoromethyl)oxirane-2,2-dicarboxylate. Negative shift values correspond to a blue shift.

The VCD of the nitro-derivative has been simulated in the gas phase using the B3LYP[38]-D4/def2-QZVPPD level, yielding a similarity score of 96% (the result of this study is shown in Supporting Information Figure S5). In all cases, the weighting function is shown to highlight the relevant areas of the spectrum and indicate the ones that have been mostly considered during the optimization process. Further validation on two additional systems with distinct spectral patterns is provided in the Supporting Information.[39]

3.4. Estimation of Singlet−Triplet Energy ($\Delta E_{S\to T}$)

Another valuable property of the molecules is the Singlet−Triplet energy gap ($\Delta E_{S\to T}$). This calculation relies on different multiplicities of the whole system. EnAn supports this calculation by allowing users to define both charge and multiplicity for each step in the protocol, enabling workflows for singlet−triplet gaps and other related properties. To demonstrate this capability, we calculated the energy gap for a recently published aminoborane[40] compound. The protocol was carried out in CPCM($CHCl_3$) and composed as (i) the optimization of the S_0 level at the r^{2}SCAN-3c level, followed by (ii) a single-point calculation at CAM-B3LYP-D4/def2-TZVPP, (iii) the optimization using ΔSCF[41,42] with HOMO−LUMO configuration of the α-spin electron for the S_1 electronic state, (iv) followed by a second single-point energy evaluation, and finally, (v) the optimization of the T_1 state, and (vi) its single-point step. A simplified Jablonski diagram summarizing the workflow is shown in Figure 4.

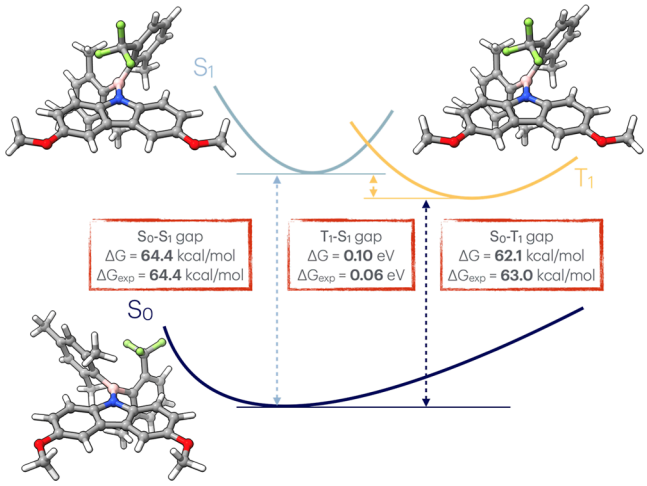


Figure 4. Simplified Jablonski diagram illustrating the singlet and triplet electronic states of a case-study aminoborane.

3.5. Transition-State Ensemble Refinement

The conformational search is critical not only at a ground-state level but also for the TSs; this analysis can lower the estimated TS energy by up to 10 kcal/mol.[10] EnAn has been designed to refine TS ensembles, as well. It is possible to insert additional inputs into the protocol definition in each different step. Thus, we created a case study involving two flexible reagents for the classic Diels−Alder reaction. First, a candidate TS geometry was located using Nudged Elastic Band (NEB)[43−45] at the GFN2-xTB level using ORCA, then optimized toward the saddle point at the r^{2}SCAN-3c level, and finally, its electronic energy was evaluated at the ωB97X-3c[28] level. From this geometry, an ensemble of possible TSs has been generated using GOAT, and this massive group of conformers (673) has been processed into EnAn with the following protocol: (i) creation of 30 clusters at the GFN2-xTB energy level, (ii) constrained optimization at r^{2}SCAN-3c freezing the four reacting atoms, (iii) the actual optimization toward the saddle point at the same level, and (iv) the final electronic energy calculation at ωB97X-3c, in order to obtain comparable energy values.

This refinement decreased the TS energy by 4.8 kcal/mol (Figure 5). This example demonstrates that even small changes

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025

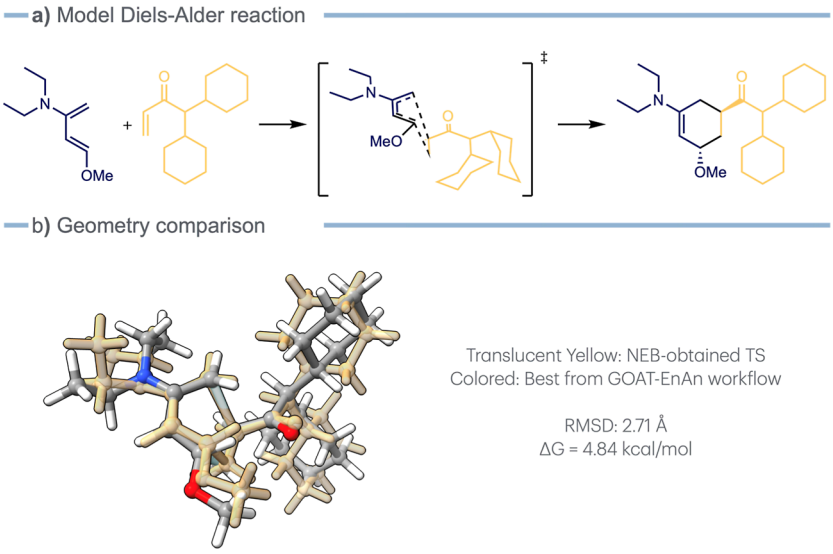


Figure 5. (a) Case study Diels–Alder reaction. (b) Comparison between the TS obtained through the NEB workflow (translucent yellow) and the one refined through the GOAT-EnAn workflow.

in some peripheral substituents of the reactants or catalyst might significantly alter reaction pathway energies, potentially affecting the selectivity of the process studied.

4. CONCLUSIONS

In this work, we present a computational framework designed to automate the management of conformational ensembles by interfacing with multiple QM engines. The sheer size of such ensembles often renders manual processing intractable, necessitating automated solutions.

A distinct advantage of this software is that it enables researchers, regardless of their programming expertise, to define reproducible text-based protocols for selecting representative conformers. This filtering process is driven by pairwise comparisons against user-defined thresholds for rotational constants and electronic energies alongside PCA-guided clustering.

The versatility of the tool was demonstrated through diverse case studies: (i) the optimization and refinement of ensembles for medium to large, highly flexible scaffolds; (ii) ensemble reduction via PCA clustering; (iii) the simulation of VCD and ECD spectra for chiral systems; (iv) the assessment of singlet–triplet energy gaps in a recently reported system; and (v) the refinement of TS ensembles for model reactions bearing flexible substituents. In all scenarios, the pipeline successfully streamlined the workflow, yielding robust results with minimal user intervention.

■ ASSOCIATED CONTENT

Data Availability Statement

The source code is available on GitHub at https://github.com/andre-cloud/ensemble_analyzer and archived at DOI: 10.5281/zenodo.18255912.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.6c00273.

Installation instructions and system requirements; theoretical background on conformational ensemble refinement and spectral property calculations; comprehensive usage guide with step-by-step tutorials; and detailed computational protocols and parameters for all case studies (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Andrea Pellegrini − *Department of Industrial Chemistry "Toso Montanari", Alma Mater Studiorum University of Bologna, 40129 Bologna, Italy;* orcid.org/0000-0001-5322-5949; Email: andrea.pellegrini15@unibo.it

Authors

Paolo Righi − *Department of Industrial Chemistry "Toso Montanari", Alma Mater Studiorum University of Bologna, 40129 Bologna, Italy;* orcid.org/0000-0003-1419-2560

Andrea Mazzanti − *Department of Industrial Chemistry "Toso Montanari", Alma Mater Studiorum University of Bologna, 40129 Bologna, Italy;* orcid.org/0000-0003-1819-8863

Michele Mancinelli − *Department of Industrial Chemistry "Toso Montanari", Alma Mater Studiorum University of Bologna, 40129 Bologna, Italy;* orcid.org/0000-0002-8499-5265

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.6c00273

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. A.P. was responsible for conceptualization, software, and investigation. A.P., P.R., A.M., and M.M. contributed to discussion and writing—review and editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the "Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali" (INSTM) and the University of Bologna for financial support. A.P. acknowledges the financial support from the National

Recovery and Resilience Plan (NRRP), Mission 4, Component 1, Investment 4.1 by the Italian Ministry of University and Research (MUR), funded by the European Union (CUP J33C23002180002). We thank Rebecca Nucci and Rita Tucci for testing the frameworks and reporting bugs.

REFERENCES

(1) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148* (24), 241733.
(2) Zahrt, A. F.; Rinehart, N. I.; Denmark, S. E. A Conformer-Dependent, Quantitative Quadrant Model. *Eur. J. Org. Chem.* **2021**, *2021* (17), 2343−2354.
(3) Bursch, M.; Mewes, J.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angew. Chem.* **2022**, *134* (42), No. e202205735.
(4) Zhou, Y.; Limbu, I.; Garson, M. J.; Krenske, E. H. Conformational Sampling in Computational Studies of Natural Products: Why Is It Important? *J. Nat. Prod.* **2024**, *87* (10), 2543−2549.
(5) Pracht, P. Conformational Pruning via the Permutation Invariant Root-Mean-Square Deviation of Atomic Positions. *J. Chem. Inf. Model.* **2025**, *65* (9), 4501−4511.
(6) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **2020**, *60* (4), 2044−2058.
(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* (14), 7169−7192.
(8) 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* (11), 114110.
(9) Talmazan, R. A.; Podewitz, M. PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent. *J. Chem. Inf. Model.* **2023**, *63* (17), 5400−5407.
(10) De Souza, B. GOAT: A Global Optimization Algorithm for Molecules and Atomic Clusters. *Angew. Chem., Int. Ed.* **2025**, *64* (18), No. e202500393.
(11) Neese, F. Software Update: The ORCA Program System—Version 5.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2022**, *12* (5), No. e1606.
(12) Neese, F. Software Update: The ORCA Program System—Version 6.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2025**, *15* (2), No. e70019.
(13) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding Quantum Chemistry Methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11* (2), No. e1493.
(14) 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* (3), 1652−1671.
(15) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A Guide to Small-Molecule Structure Assignment through Computation of (1H and 13C) NMR Chemical Shifts. *Nat. Protoc.* **2014**, *9* (3), 643−660.
(16) 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* (19), 4039−4054.
(17) Alegre-Requena, J. V.; Sowndarya S V, S.; Pérez-Soto, R.; Alturaifi, T. M.; Paton, R. S. AQME: Automated Quantum Mechanical Environments for Researchers and Educators. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2023**, *13* (5), No. e1663.
(18) Young, T. A.; Silcock, J. J.; Sterling, A. J.; Duarte, F. autodE: Automated Calculation of Reaction Energy Profiles—Application to Organic and Organometallic Reactions. *Angew. Chem., Int. Ed.* **2021**, *60* (8), 4266−4274.
(19) Cormanich, R. A.; Da Silva, G. D. Autobench V1.0: Benchmarking Automation for Electronic Structure Calculations. *J. Chem. Inf. Model.* **2024**, *64* (8), 3322−3331.
(20) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825.
(21) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29* (27), 273002.
(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*. Revision A.03; Gaussian Inc, 2016.
(23) Pracht, P.; Grimme, S. Calculation of Absolute Molecular Entropies and Heat Capacities Made Simple. *Chem. Sci.* **2021**, *12* (19), 6551−6568.
(24) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Trans. Math. Softw.* **1997**, *23* (4), 550−560.
(25) Froitzheim, T.; Müller, M.; Hansen, A.; Grimme, S. g-xTB: A General-Purpose Extended Tight-Binding Electronic Structure Method For the Elements H to Lr (Z = 1−103). *ChemRxiv* **2025**, *24*, bjxvt.
(26) Grimme, S.; Hansen, A.; Ehlert, S.; Mewes, J.-M. r^2SCAN-3c: A "Swiss Army Knife" Composite Electronic-Structure Method. *J. Chem. Phys.* **2021**, *154* (6), 064103.
(27) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom−Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615.
(28) Müller, M.; Hansen, A.; Grimme, S. ω B97X-3c: A Composite Range-Separated Hybrid DFT Method with a Molecule-Optimized Polarized Valence Double-ζ Basis Set. *J. Chem. Phys.* **2023**, *158* (1), 014103.
(29) 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* (18), 3297.
(30) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; Van Der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.·.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; Van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.;

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025

Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; De Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261−272.

(31) Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Bringmann, G. SpecDis: Quantifying the Comparison of Calculated and Experimental Electronic Circular Dichroism Spectra. *Chirality* **2013**, *25* (4), 243−249.

(32) Covington, C. L.; Polavarapu, P. L. Similarity in Dissymmetry Factor Spectra: A Quantitative Measure of Comparison between Experimental and Predicted Vibrational Circular Dichroism. *J. Phys. Chem. A* **2013**, *117* (16), 3377−3386.

(33) Battaglia, V.; Sonsona, I. G.; Pellegrini, A.; Mazzanti, A.; Lattanzi, A. Installation of Trifluoromethylated Quaternary Carbon Stereocenters via Asymmetric Epoxidation of Tetrasubstituted Alkenes. *Adv. Synth. Catal.* **2025**, *367*, No. e70153.

(34) Martinelli, E.; Vicini, A. C.; Mancinelli, M.; Mazzanti, A.; Zani, P.; Bernardi, L.; Fochi, M. Catalytic Highly Enantioselective Transfer Hydrogenation of *β*-Trifluoromethyl Nitroalkenes. An Easy and General Entry to Optically Active *β*-Trifluoromethyl Amines. *Chem. Commun.* **2015**, *51* (4), 658−660.

(35) 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* (18), 6378−6396.

(36) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange−Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393* (1−3), 51−57.

(37) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and Evaluation of the D4 London-Dispersion Model for Periodic Systems. *Phys. Chem. Chem. Phys.* **2020**, *22* (16), 8499−8512.

(38) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98* (45), 11623−11627.

(39) Monda, G.; Kiriakidi, S.; Nieto Faza, O.; Mazzanti, A.; Bertuzzi, G.; Bandini, M. Enantioselective Pd-Catalyzed Electrochemical Dearomative Allylation of Tropones: Construction of All-C Quaternary Stereocenters. *Org. Lett.* **2026**, *28*, 4227.

(40) Matera, N.; Bussolari, A.; Mancinelli, M.; Coccia, R.; Santarelli, N.; Pellegrini, A.; Mazzanti, A. Aminoboranes as Rationally Tuned Organic Photosensitizers for Energy Transfer Catalysis. *Angew. Chem., Int. Ed.* **2026**, *65*, No. e20339.

(41) Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). *J. Phys. Chem. A* **2008**, *112* (50), 13164−13171.

(42) Hait, D.; Head-Gordon, M. Orbital Optimized Density Functional Theory for Electronic Excited States. *J. Phys. Chem. Lett.* **2021**, *12* (19), 4517−4529.

(43) Mills, G.; Jónsson, H.; Schenter, G. K. Reversible Work Transition State Theory: Application to Dissociative Adsorption of Hydrogen. *Surf. Sci.* **1995**, *324* (2−3), 305−337.

(44) Berne, B. J.; Ciccotti, G.; Coker, D. F. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific, 1998.

(45) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113* (22), 9978−9985.

https://doi.org/10.1021/acs.jcim.6c00273
J. Chem. Inf. Model. 2026, 66, 5018−5025