Molecular Gas Phase Conformational Ensembles

Susanta Das, Kenneth M. Merz, Jr.*

Department of Chemistry, Michigan State University, 578 S. Shaw Lane, East Lansing, Michigan 48824, United States

*E-mail: merzjrke@msu.edu

Abstract. Accurately determining the global minima of a molecular structure is important in diverse scientific fields, including drug design, materials science, and chemical synthesis. Conformational search engines serve as valuable tools for exploring the extensive conformational space of molecules and identifying energetically favorable conformations. In this study, we present a comparison of Auto3D, CREST, Balloon, and ETKDG (from RDKit), which are freely available conformational search engines, to evaluate their effectiveness in locating the global minima. These engines employ distinct methodologies, including machine learning (ML) potential-based, semiempirical, and force field (FF) based approaches. To validate these methods, we propose the use of collisional cross section (CCS) values obtained from ion mobility – mass spectrometry (IM-MS) studies. We hypothesize that experimental gas-phase CCS values can provide experimental evidence that we likely have the global minimum for a given molecule. To facilitate this effort, we used our gas-phase conformation library (GPCL) which currently consists of the full ensembles of 20 small molecules, which can be used by the community to validate any conformational search engine. Further members of the GPCL can be readily created for any molecule of interest using our standard workflow used to compute CCS values expanding the ability of the GPCL in validation exercises. These innovative validation techniques enhance our understanding of the conformational landscape and provide valuable insights into the performance of conformation generation engines. Our findings shed light on the strengths and limitations of each search engine, enabling informed decisions for their utilization in various scientific fields, where accurate molecular structure determination is crucial for understanding biological activity and designing targeted interventions. By facilitating the identification of reliable conformations, this study significantly contributes to enhancing the efficiency and accuracy of molecular structure determination, with a particular focus on metabolite structure elucidation. The findings of this research also provide

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

valuable insights for developing effective workflows in predicting the structures of unknown compounds with high precision.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Introduction:

Accurately determining molecular ensembles is crucial in computational chemistry for understanding molecular behavior and properties. However, predicting the global minima, which represents the most stable conformation, presents a significant challenge due to the size of conformational spaces for flexible molecules. Moreover, to correctly rank order all low energy conformations poses another significant technical challenge. To overcome these issues, efficient and reliable conformational search algorithms are necessary to explore this space. Conformer generation plays a pivotal role in various computational analyses, including computational drug design[1,2], 3D QSAR modeling[3,4,] protein-ligand docking[5–8], and structure elucidation of unknown compounds[9–11]. Different methods exist for generating conformers, ranging from obtaining a single low-energy conformation to generating ensembles that encompass biologically relevant low-energy conformational space. The choice of conformational sampling technique directly influences the subsequent analysis's reliability and speed.

Multiple conformational search engines are available including, for example, Balloon[12,13], ETKDG from RDKit[14–16], Confab[17], Frog2[7,18], MacroModel[19,20], OMEGA[21,22], CREST[23–25], and Auto3D[26]. These tools offer diverse methods and algorithms for conformation generation, ranging from force field-based approaches to semiempirical and machine learning potential-based methods. Various software tools such as Balloon, ETKDG, Confab, Frog2, and MacroModel employ force field-based methods to generate conformation ensembles, combining systematic and random sampling techniques within the framework of a force field to explore molecular conformational space. Balloon, for instance, leverages a blend of systematic and random sampling techniques, allowing broad coverage of molecular conformations, encompassing both low-energy and higher-energy regions.

This comprehensive exploration of conformational space gives a more complete picture of the conformational landscape. RDKit, an open-source cheminformatics toolkit, employs a distance geometry algorithm (ETKDG) along with distance constraints derived from a force field.[16,27,28] Confab, provided by the Molecular Operating Environment (MOE) software package, integrates systematic and random sampling methods within a force field framework.[29–35] Frog2, developed by Certara, utilizes a proprietary algorithm based on force field methods to sample low-energy conformations of drug-like molecules.[7,13,18,36–42] MacroModel, offered by Schrödinger, employs molecular mechanics force fields such as OPLS-AA to explore conformational space.[43–52] Omega, developed by OpenEye Scientific Software, is a conformation generation tool that combines distance geometry, systematic search, and random perturbation methods to generate diverse conformations.[53–63] CREST (Conformer-Rotamer Ensemble Sampling Tool)

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

utilizes the semiempirical GFN2-xTB model, which is a parametrized self-consistent tight-binding quantum chemical method using multipole electrostatics and density-dependent dispersion contributions to calculate energy profiles and explore conformational space.[64–70] Auto3D employs machine learning potential-based methods, utilizing an artificial neural networks trained on a large dataset of molecular structures to predict energetically favorable conformations.[71–79]

In addition to these engines, there are several other notable conformation generation tools available. The BioChemical Library, BCL::Conf is a conformational sampling tool developed by Meiler *et. al* that utilizes a combination of systematic search, stochastic optimization, and diversity analysis methods.[80] The Experimental-Torsion Distance Geometry with basic Knowledge (ETKDG), which is offered via RDKit, is a stochastic search method that uses distance geometry and knowledge from experimental crystal structures to explore the conformational space.[81] Conformator is a conformation search engine provided by the NAOMI ChemBio Suite that generates conformer ensembles using an incremental construction approach.[82] The CSD conformer generator is a tool specifically designed for generating conformations of small organic molecules using information from the Cambridge Structural Database (CSD).[83] ConfGen, developed by Schrödinger, is a conformation generation tool that combines systematic search and molecular dynamics simulations to explore conformational space.[61] CORINA is a conformational search tool that utilizes a combination of stochastic search algorithms, distance geometry, and energy minimization to generate low-energy conformations.[84] MOE (Molecular Operating Environment), a software package from Chemical Computing Group, provides a suite of conformational search algorithms and methods for generating conformational ensembles.[85] Other conformation generation engines include iCon, which employs an incremental construction approach to systematically explore the conformational space[72], and CAESAR, a tool that combines genetic algorithms with energy minimization to generate low-energy conformations.[54]

Apart from the aforementioned conformation generation tools, there exists a diverse range of algorithms specifically designed for generating conformer ensembles. These algorithms facilitate conformational sampling in both gas and solution phases, allowing for a more thorough exploration of molecular flexibility. These include Confort[86], ROTATE[87], CONFECT[88], Catalyst[89,90], MED-3DMC[91], Multiconf-DOCK[92], CONFECT[93], BRIKARD[94], ForceGen[95], TCG (TrixX Conformer Generator)[96] and Cxcalc (ChemAxon)[97]. These tools utilize a range of algorithms and methodologies to explore the conformational space of molecules and generate conformational ensembles. ROTATE employs a systematic search algorithm based on molecular flexibility, while Catalyst utilizes a stochastic

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

search algorithm with a focus on energy optimization. Confort incorporates a distance geometry approach to generate low-energy conformations, while MED-3DMC utilizes a Monte Carlo-based method. Multiconf-DOCK utilizes a systematic search approach for exploring ligand flexibility within the DOCK5 program. It extends multiple anchor segments stepwise and generates conformations by systematically rotating single, nonterminal, acyclic bonds at specified increments, while CONFECT employs an evolutionary algorithm. BRIKARD utilizes a knowledge-based approach, and ForceGen incorporates force field-based methods. TCG utilizes a systematic torsion angle search algorithm and Cxcalc utilizes a fragment fusion method and the Dreiding force field for the calculation and optimization of conformers. These tools aid in the exploration of potential binding modes and interactions.[97,98] Finally, it's important to note that all of these methods generate conformations in the gas-phase and not solution or in the crystalline phase.

The primary objective of these conformation generation tools is to identify the global minima or a list of low-energy conformers from a large ensemble of generated conformations. The accuracy, speed, and computational reliability of these tools are achieved through different algorithmic approaches.[99,100] However, it is crucial to validate the results obtained from these tools with experimental findings. The validation of ligand conformations often involves comparing the generated conformers with experimentally determined structures, typically obtained from protein-bound ligand conformations extracted from the Protein Data Bank (PDB).[101–110] A shortcoming of these so-called "bioactive conformers" for the validation of conformation generation software is the limited number and diversity of experimentally determined protein-ligand structures and questions surrounding whether these "bioactive" conformers represent the global minimum or local minimum or high energy structures in the conformational ensemble.[84,111–126] Additionally, X-ray structures in the PDB represent static snapshots of molecules in crystalline states, which may not fully capture their dynamic behavior in solution or other environments. The resolution of X-ray structures is used as a quality criterion and low-resolution structures may lack precision and atomic-level details necessary for accurate conformation determination. It is crucial to consider these limitations and explore alternative validation approaches, such as benchmark datasets or comparison with other experimental data.[101–103,127–130] In addition to the Protein Data Bank (PDB), another widely used validation dataset for ligand conformations is the Cambridge Structural Database (CSD).[131] The CSD primarily consists of small organic molecule crystal structures obtained from X-ray crystallography experiments. It offers a large collection of experimentally determined structures, providing valuable insights into the three-dimensional arrangements and intermolecular interactions of small molecules in the solid state.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

The use of the CSD as a validation dataset complements the information obtained from the PDB, expanding the scope of ligand conformation validation and contributing to a more comprehensive understanding of ligand behavior in different environments.

Over the past two decades, Ion Mobility-Mass Spectrometry (IM-MS)[132] has greatly advanced as an important method in metabolomics applications for the characterization of small molecules in the gas-phase. The technique of ion mobility spectrometry (IMS) measures an ions drift time through a gas-filled region (N_2 or He) under the influence of an electric field, efficiently separating gas-phase molecular ions based on charge, size, and shape. IM-MS integrates diverse technologies like traveling wave (TWIMS)[133], drift tube (DTIM)[134], trapped IM (TIMS)[135], and differential mobility MS (DMS)[136], in combination with ionization methods such as matrix-assisted laser desorption ionization (MALDI), electrospray ionization (ESI), and atmospheric pressure chemical ionization (APCI). It seamlessly accommodates liquid-phase separations including capillary electrophoresis, gas chromatography, and supercritical fluid chromatography. Unlike traditional techniques like X-ray crystallography and NMR spectroscopy, IM-MS is rapid, bypassing the need for prior purification or crystallization of the target compound. The technique doesn't directly measure molecular surfaces, instead, it derives collision cross-section (CCS) values from mobility data using a mathematical model. These CCS values are key for understanding the conformational features of molecules and enable comparison with experimental values for accurate and consistent results. The term Collision Cross Section was historically associated with the hard sphere collision model.[137] Theoretically, CCS is computed by averaging the cross-sectional area of rotation for a target molecular ion using the Mason-Schamp equation.[138,139] This process relies on input atomic coordinate files derived from various sources, including X-ray scattering, NMR data, quantum chemical calculations, or molecular dynamics (MD) simulations. Three distinct computational treatments of ion-buffer gas collisions are in general use: projection approximation (PA)[140], exact hard sphere scattering (EHSS)[141], and the trajectory method (TM)[141] are all employed in CCS estimation. To achieve accurate CCS predictions, it is essential to calculate all potential collision angles between a buffer gas and the target molecule. The precision of CCS prediction is contingent upon effectively sampling the correct conformational space. Consequently, the initial generation of the conformation ensemble plays a pivotal role in predicting accurate CCS values and, by extension, the correct three-dimensional structure in the gas phase.

Hence, in this work, we hypothesize that gas-phase IM-MS studies sample the gas-phase conformational ensemble of a molecule providing unique insights into the shape and, hence, the structure of a molecule in the gas-

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

phase. Importantly, we are assuming the drift gas (usually N_2 or He) has a minimal perturbation on the gas-phase molecular structure. Based on this hypothesis, we propose a novel approach to evaluate and compare gas-phase conformational search engines based on their ability to characterize the gas-phase conformational ensemble and identify the global minima using a quantum mechanics (QM) based workflow whose outcomes are compared against experimental IM-MS information.[9,10] Specifically, we have developed a QM-based method to calculate Collisional Cross Sections (CCS), which is an accurate indicator of the global minima for molecular structures in the gas-phase[10]. For example, we have shown[10] (*vide infra*) that if you take a higher energy, non-global minimum and compute its CCS value the deviation from the experimental CCS values is increased. Our CCS calculations have been validated against experimental data, demonstrating their reliability in capturing the most stable conformations.[142–147] To conduct our comparative analysis, we employed four different freely available conformational search engines: Auto3D, CREST, Balloon, and ETKDG (from RDKit). These engines utilize diverse methodologies, including force field-based conformation generation (ETKDG, Balloon), semi-empirical methods (CREST), and machine learning potentials (Auto3D). By generating conformations using each engine and comparing them with the ensemble and global minima validated through CCS calculations, we aimed to identify the most effective conformational search engine for accurate global minima prediction. By evaluating and comparing the performance of different engines in the gas phase, our study aims to provide valuable insights into the selection of the optimal conformational search approach for improved molecular structure determination and related applications. Moreover, the resultant data set can be used to validate other gas-phase conformational search engines.

Computational Methods:

In this study, we focused on 20 metabolites and employed a DFT based workflow to compute their Collisional Cross Section values. Our workflow has demonstrated good accuracy in CCS prediction, with an error rate of less than 3% compared to experiment (experimental error is ~3%). Our established workflow encompasses the following steps to predict accurate CCS values: First, the conformations of each metabolite were generated using tools contained within RDKit.[14,15] The conformation of the molecule is initially generated using the ETKDG algorithm, an embedded method within RDKit, employing default settings, followed by structural optimization of the generated conformations using the MMFF94 force field. A maximum number of generated conformers is set to 1000 for the small molecules systems. Each generated conformer was then geometry optimized using the ANI QM-ML model.[74,76,148] The optimized structures were subsequently clustered using our in-house automated clustering code

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

AutoGraph, enabling the identification of chemically distinct conformations as centroids of the identified clusters.[149–151] Geometry optimization and Mulliken atomic charge calculations were performed on representative conformations of each identified cluster using B3LYP/6-31+G(d,p) and B3LYP/6-311++G(d,p) level of theory, respectively employing a GPU enabled, in-house developed QM engine called QUICK.[152–154] The CCS values were computed using the trajectory method (TM) as implemented in the HPCCS code developed by Zanotto et al.[155,156] The inclusion of an unsupervised clustering method in our workflow reduces the potential for human bias and error in cluster selection, while the QM-ML model and clustering technique contribute to its computational efficiency.

To assess the accuracy of conformational search engines in predicting the global minima, we compared the generated conformations from Auto3D, CREST, Balloon, and ETKDG (from RDKit), with the most stable conformation determined by our QM based workflow. Conformations were ranked based on increasing relative energies, computed using the respective potential energy functions employed by each conformation generation tool. We performed RMSD calculations between the generated conformations and the QM optimized most stable conformation using the LS-align algorithm, a high-throughput virtual screening atom-level structural alignment method developed by Zhang *et al*.[157] The conformation with the lowest RMSD and energy values was considered the global minima for that particular molecule using the specific conformation generation engine. If no conformation matched these criteria, it was deemed that the engine failed to find the global minima for that molecule.

Furthermore, we calculated the Boltzmann average CCS values using the conformations generated by the conformation search engines and compared them with experimental values. The percentage error in predicting the CCS was reported and an error range within ± 3% was considered indicative of a good CCS prediction as the experimental uncertainty of CCS values within ± 3%.

System setup. In our previous study, we extensively investigated various ionization models (protonation/deprotonation) and their impact on CCS prediction accuracy for metabolites.[10] The predicted CCS values were compared to experimental results to identify the charge model that exhibited the lowest error percentage. In the current study, focusing on finding global minima based on CCS values, we selected the protonation state that yielded the best predicted CCS values (lowest error percentage) for further analysis. For instance, in the case of the carnosine molecule, five models were considered (model 1, model 2, model 3, model 4, and model 5) with corresponding CCS errors of 9.4%, 9.9%, 8.3%, 0.1%, and 31.0%, respectively. Model 4, exhibiting the lowest error percentage, was

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

chosen as the representative carnosine model for the present investigation. Figure 1 presents an overview of the metabolites included, with their respective ionization sites. The protonation site is distinctly highlighted in red, while the deprotonation site is accentuated in blue.

In the context of open-source applications, we used the default settings provided by the developers of the methods. In the case of ETKDG (from RDKit), the conformations were generated and then optimized using the MMFF94 force field. The 'rdkit.Chem.AllChem' module was utilized to generate conformations through a Python script. The number of generated conformations was set to 1000 by invoking the 'EmbedMultipleConfs' keyword, with the RMS threshold set to 0.1 kcal mol^{-1} $Å^{-1}$ (pruneRmsThresh = 0.1). The 'randomSeed' keyword was employed to acquire a seed for the random number generator. Following this, the generated conformations underwent optimization using the MMFF94 force field, and the resulting optimized coordinates and energy were stored for subsequent analysis. Balloon also uses energy minimization employing the MMFF94 (file name- 'MMFF94.mff') force field, which is defined using the '-forcefield' ('-f') setting in Balloon, and the input and output file formats are in SDF. Balloon employed a multiobjective genetic algorithm for generating ensembles of molecular conformers. The number of generated conformations was determined by the 'nconfs' switch, which was set to 1000, and the convergence criterion based on the gradient root-mean-square (rms) was set to 0.1 kcal mol^{-1} $Å^{-1}$. Balloon conducts conformation generation, RMSD checks, and eliminates similar conformations as described in the original publication.[13] It is noteworthy that, despite the initial ensemble size being set at 1000 conformations, the ultimate ensemble size varied based on the number of rotatable bonds.. For CREST, conformer generation employs the GFN2-xTB method.[73] CREST utilizes metadynamics[158] for conformation generation and writes input and output files in the xyz file format. The 'ewin' keyword sets the default energy window for conformational energies at GFN2-xTB level in CREST, typically at 6 kcal/mol. The 'cluster' keyword is used to cluster the generated conformations, while the 'chrg' keyword designates the molecule's charge. Auto3D initially runs the default RDKit conformer engine ETKDG using SMILES as input for stereoisomer enumeration and 3D construction. The input command 'K=1000' generated 1000 conformations, which were subsequently filtered using a 0.3 Å RMSD constraint to enhance structural diversity. Conformations are optimized by the 3D optimization engine of AIMNET[159,160], which locates the geometry corresponding to a stationary point on the potential energy surface by computing energy and analytic forces in each optimization step. AIMNET can manage neutral and charged molecules. OpenBabel's root mean square deviation

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(RMSD) filters duplicate conformers. The structures are ranked by energies and duplicates are removed, consolidating low-energy structures into a single SDF file.

Results and Discussions:

Conformation generation and global minima search. In this study, we examined the performance of various conformation generation engines, including Auto3D, CREST, Balloon, and ETKDG (from RDKit), in generating conformations and identifying global minima. Table 1 provides an overview of the number of conformations generated by each engine for the selected metabolites. In the case of QM results, the conformation generation process involved using ETKDG to initially generate conformations, followed by clustering and subsequent QM geometry optimization.

Among the engines, Auto3D and CREST had the capability to perform clustering as part of their conformation generation process, whereas Balloon and ETKDG did not include this clustering step. Consequently, after generating conformations using Balloon and ETKDG, we applied the Autograph clustering algorithm to cluster the resulting conformational ensemble. This allowed for an analysis of the conformations and their subsequent evaluation in terms of capturing the global minima. The number of conformations generated by Balloon and ETKDG prior to the clustering step can be found in Table S1-S64 in the SI. The inclusion of a clustering step in Auto3D and CREST eliminated high energy conformations giving a short list of conformations to consider. On the other hand, Balloon and ETKDG produced a significantly higher number of conformations due to the lack of this pruning step. It is worth noting that the number of conformations generated by Balloon was lower than that of ETKDG.

To determine the global minima for each molecule, we employed a root-mean-square deviation (RMSD) matrix to compare the conformations generated by the conformation search engines with the lowest-energy QM conformation. The RMSD values for all conformations can be found in the SI specifically Table S1-S102. The conformations were ranked based on both RMSD and energy values, and those achieving the top rank (ranked as 1, lowest RMSD, lowest energy) in both categories were considered global minima and are highlighted in **bold** in Table 2. On the other hand, if the lowest-energy conformation did not correspond to the lowest RMSD value, it indicated that the engines failed to identify the global minima, and these instances are not highlighted. For instance, in the case of carnosine, Auto3D successfully identified the global minima with a rank of 1 out of 13 conformations, while ETKDG achieved a rank of 1 out of 9 conformations. However, CREST and Balloon were unable to find the global minima, as their lowest RMSD conformations ranked 7 out of 7 and 6 out of 10 in terms of energy, respectively. The

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

details of the RMSD values, relative energies, and corresponding rank of Carnosine conformations are given in Table 3. It is important to note that the range of relative energies obtained from the conformation generation engines exhibits significant variation. Our analysis revealed that the relative energies generated by Auto3D span a wide range, while the relative energies produced by CREST are relatively compressed. For the carnosine system, all seven conformations generated by CREST exhibited relative energies within 5 kcal/mol, whereas none of the 13 conformations generated by Auto3D fell within this range. Notably, Conformation 9 (*viz.* Conf_9) was the second lowest in energy among the Auto3D conformations, but it had a higher energy by 15.9 kcal/mol. The highest relative energies obtained from Balloon and ETKDG were 6.3 kcal/mol and 17.8 kcal/mol, respectively. In comparison, the highest energy conformations generated by Auto3D and CREST were 31.9 kcal/mol and 5.7 kcal/mol, respectively. Detailed energy values for all the molecules can be found in the Supporting Information (Table S1-S102). Out of the 20 metabolites considered in this study, Auto3D successfully identified the global minima for 8 metabolites, whereas CREST detected them for 4 metabolites. Balloon and ETKDG each demonstrated success for 3 metabolites. The success rate of Auto3D in finding global minima was 40%, while CREST attained 20%. Balloon and ETKDG both achieved a 15% success rate, indicating comparatively lower performance.

CCS Prediction and Comparison. We further evaluated the accuracy of CCS predictions for the generated conformational ensembles using all the engines, as summarized in Table 4. The calculated CCS values were compared with experimental CCS values, and predictions within 3% of the experimental values were considered accurate and bolded. Conversely, predictions with errors exceeding 3% were considered inaccurate and not highlighted.

Our results showed that out of the 20 metabolites, the QM method achieved accurate CCS predictions for 13 metabolites, resulting in a success rate of 65%. Among the conformation generation engines, Auto3D demonstrated accurate CCS predictions for 8 molecules, yielding a success rate of 40%. CREST performed well, achieving accurate CCS predictions for 9 metabolites with a success rate of 50%. However, Balloon and ETKDG exhibited lower accuracy, correctly predicting CCS values for only 5 and 6 metabolites, respectively, with success rates of 25% and 30%. The average errors in CCS predictions for the QM, Auto3D, CREST, Balloon, and ETKDG generated conformational ensembles were found to be 2.5%, 4.9%, 3.7%, 5.9%, and 6.0%, respectively. Notably, the QM method achieved the highest accuracy in CCS prediction, highlighting its superiority in capturing the conformational behavior of the metabolites. The semi-empirical-based engine CREST had a success rate of 50% in accurately predicting CCS

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

values. However, the ML-based engine Auto3D exhibited a slightly lower accuracy rate of 40%. The force field-based engines Balloon and ETKDG yielded the lowest accuracy rates of 25% and 30%, respectively.

These results underscore the importance of selecting appropriate conformation generation engines for accurate prediction of global minima and CCS values in molecular gas phase conformational ensembles. The QM method, with its ability to capture fine structural details and accurately calculate CCS values, emerges as the most reliable approach. The findings also highlight the promising performance of the semi empirical based engine CREST and the ML based engine Auto3D, while indicating the limitations of FF based tools such as Balloon and ETKDG in accurately representing the conformational space and predicting CCS values.

Gas Phase Conformational Library (GPCL). The establishment of a Gas Phase Conformational Library (GPCL) is a steps towards developing a data set for the validation of gas-phase conformational search engines. GPCL, currently encompasses the full ensembles of 20 small molecules. This library, constructed through our standard QM workflow to compute CCS values, is a freely available dataset for validating various conformational search engines. Moreover, the data set can be used to refine conformational generation methods, thereby enhancing the reliability of computational predictions. While the GPCL only contains 20 molecules it can be readily expanded using our standard workflow to generate additional members of any molecule of interest, significantly expanding the library's utility. Because the GPCL is open source it forms the basis of a collaborative platform that can be used to help develop and validate conformational search engines.

Conclusion:

In this study, we investigated the performance of different conformation generation engines, namely Auto3D, CREST, Balloon, and ETKDG (from RDKit), in generating molecular gas phase conformational ensembles by their ability to predict experimental gas-phase collisional cross section values for 20 small molecules. We utilized a computational workflow that encompassed conformer generation, clustering, and analysis of global minima and accurate CCS prediction. The conformations were generated using the respective conformational search engines, and we compared them with the global minima obtained through QM computation that were validated against experimental CCS values. We also compared the predicted CCS values of the generated conformations with experimental values to assess their accuracy. Based on this analysis, we observed that the ML based algorithm, Auto3D achieved the highest

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

success rate in identifying global minima, followed by CREST, ETKDG, and Balloon. In terms of CCS prediction accuracy, QM methods yielded the most accurate results, followed by CREST, Auto3D, ETKDG and Balloon. It is noteworthy that while Auto3D demonstrated a higher success rate in global minima identification, CREST exhibited relatively higher accuracy in CCS prediction among the engines considered.

This study provides insights into the performance of different conformation generation engines and their impact on global minima identification and accurate CCS prediction. Moreover, the findings will contribute to the development of more reliable computational workflows for conformational search and related applications in drug design. Based on our present observations conformational search tools have significant room for improvement for gas-phase ensemble prediction. To help in fostering improvements we created the open-source GPCL database, which currently contains the conformational ensembles of 20 small molecules.

Our study emphasizes the significance of selecting appropriate conformation generation engines for the accurate prediction of molecular gas phase conformational ensembles, which has broad implications for drug design and metabolite structure prediction.

ASSOCIATED CONTENT

Data Availability Statement. The data associated with this study are available at a Zenodo repository (10.5281/zenodo.8356471). The repository includes all the computational results. Researchers can access and download the data to reproduce or analyze the results presented in this manuscript.

Supporting Information. The supporting information (SI) provides details on each molecule's conformation count, as well as their associated energies, RMSD values, and respective rankings based on energy and RMSD. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author. merzjrke@msu.edu [Kenneth M. Merz Jr.]

Funding Sources. The author(s) disclosed receipt of the following financial support for the research, authorship, and publication of this article: NIH (grant 1U2CES030167-01).

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Acknowledgments. The authors thank the high-performance computing center (HPCC) at Michigan State University for providing computational resources. SD and KMM acknowledge support from NIH 1U2CES030167-01.

ORCID

Kenneth M. Merz: 0000-0001-9139-5893

Susanta Das: 0000-0001-7981-5162

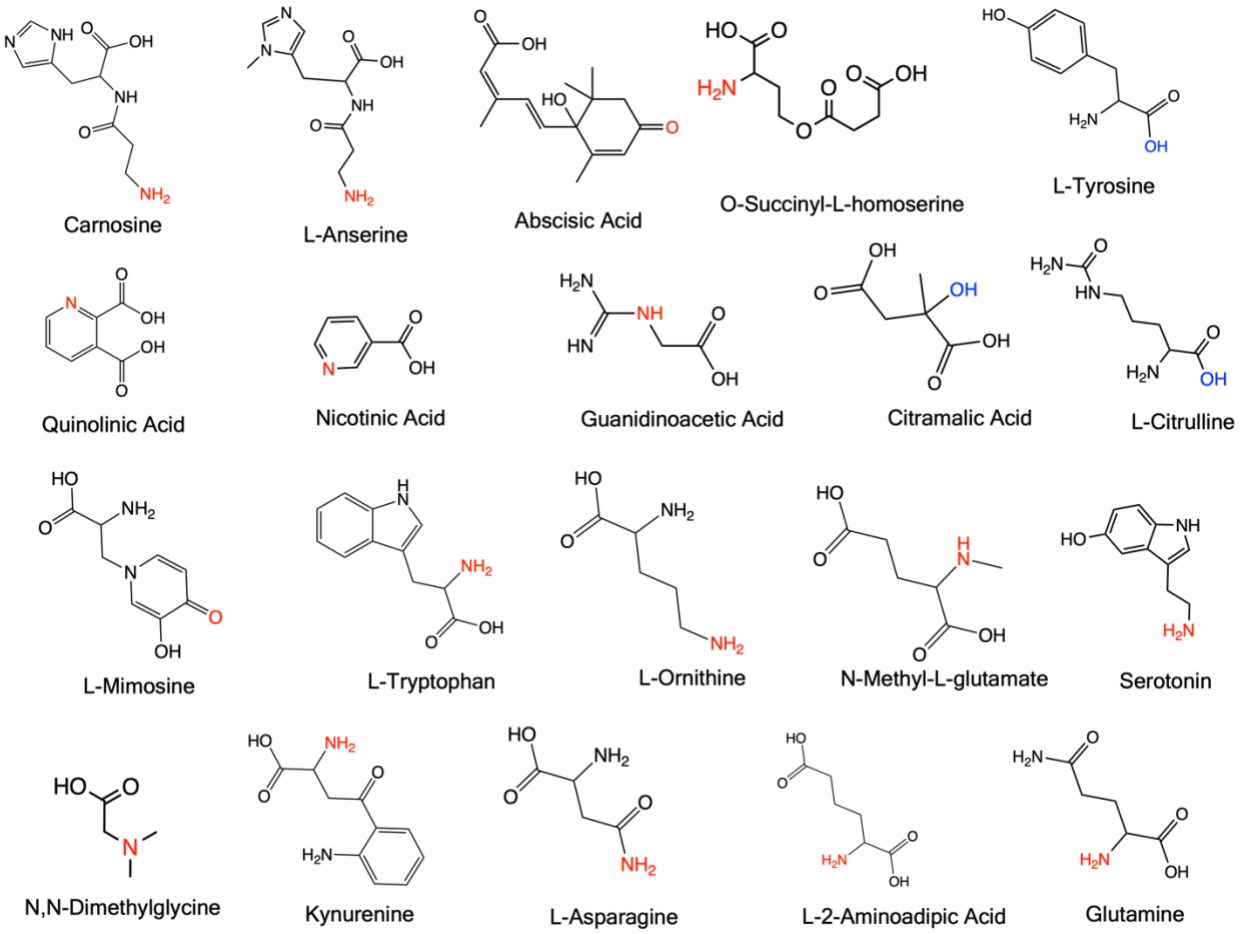


Figure 1. Metabolites examined in the study, with the protonation site highlighted in red and the deprotonation site highlighted in blue.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Table 1. The number of rotatable bonds (N.R.B) and the number of conformations identified in each case for 20 metabolites. (*The QM geometry optimized conformations after clustering using the standard workflow.)

Number	Metabolites	N.R.B	QM*	Auto3D	CREST	Balloon	ETKDG
1	Carnosine	6	12	13	7	10	9
2	O-succinyl-L-homoserine	8	17	13	16	14	23
3	L-tyrosine	3	10	5	11	3	10
4	L-mimosine	3	9	7	15	5	8
5	Citramalic acid	3	11	8	1	3	10
6	N-methyl-L-glutamate	5	15	9	5	12	13
7	L-ornithine	4	13	5	4	4	12
8	Abscisic Acid	3	10	13	9	7	10
9	L-tryptophan	3	9	9	6	4	9
10	L-asparagine	3	7	7	2	4	8
11	L-anserine	6	10	10	4	9	10
12	Kynurenine	4	8	8	2	13	12
13	Serotonin	2	4	6	5	2	7
14	N,N-Dimethylglycine	2	6	6	3	1	9
15	L-citrulline	5	17	11	17	5	14
16	Glutamine	4	9	6	6	9	9
17	L-2-Aminoadipic Acid	5	12	8	8	12	13
18	Guanidinoacetic Acid	2	10	6	4	4	11
19	Nicotinic Acid	1	3	2	3	1	4
20	Quinolinic Acid	2	5	4	3	3	5

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Table 2. Ability of Auto3D, CREST, Balloon, and ETKDG to match the lowest energy QM structure (global minimum).[a]

Number	Metabolites	Auto3D	CREST	Balloon	ETKDG
1	Carnosine	**YES (1/13)**[b]	NO (7/7)	NO (6/10)	**YES (1/9)**
2	O-succinyl-L-homoserine	NO (8/13)	NO (6/16)	NO (3/14)	NO (20/23)
3	L-tyrosine	**YES (1/5)**	NO (8/11)	NO (3/3)	**YES (1/10)**
4	L-mimosine	**YES (1/7)**	**YES (1/15)**	NO (3/5)	NO (2/8)
5	Citramalic acid	NO (6/8)	**YES (1/1)**	**YES (1/3)**	NO (6/10)
6	N-methyl-L-glutamate	NO (7/9)	NO (3/5)	NO (10/12)	NO (8/13)
7	L-ornithine	**YES (1/5)**	NO (4/4)	NO (2/4)	NO (11/12)
8	Abscisic Acid	NO (5/13)	NO (6/9)	NO (7/7)	NO (8/10)
9	L-tryptophan	NO (2/9)	**YES (1/6)**	NO (3/4)	NO (3/9)
10	L-asparagine	NO (2/7)	NO (2/2)	NO (3/4)	NO (2/8)
11	L-anserine	NO (4/10)	NO (4/4)	NO (4/9)	NO (3/10)
12	Kynurenine	**YES (1/8)**	NO (3/7)	NO (10/13)	NO (4/12)
13	Serotonin	NO (3/6)	NO (2/5)	NO (2/2)	NO (2/7)
14	N,N-Dimethylglycine	**YES (1/6)**	NO (2/3)	**YES (1/1)**	NO (6/9)
15	L-citrulline	NO (4/10)	NO (3/13)	NO (3/5)	NO (10/14)
16	Glutamine	NO (3/6)	NO (3/6)	NO (6/9)	NO (4/9)
17	L-2-Aminoadipic Acid	NO (5/8)	NO (3/8)	NO (8/12)	NO (4/13)
18	Guanidinoacetic Acid	NO (4/6)	NO (2/4)	NO (2/4)	NO (7/11)
19	Nicotinic Acid	**YES (1/2)**	NO (3/3)	**YES (1/1)**	**YES (1/4)**
20	Quinolinic Acid	**YES (1/4)**	**YES (1/3)**	NO (3/3)	NO (2/5)

a) The success or failure of Auto3D, CREST, Balloon, and RDKit-is given in columns 4, 5, 6, and 7, respectively. The bold 'YES' signifies the successful identification of global minima, while 'NO' indicates failure for a specific conformation generation engine.

b) The first number in brackets indicates the relative energy (RE) position of the conformer that best matches (in terms of RMSD) the lowest energy QM structure, while the second number is the total number of conformations in the computed ensemble. A rank of 1 (success) means the lowest energy ensemble member matches (in terms of RMSD) the lowest energy QM structure, while a rank of 2 or above (failure) means that the identified lowest energy conformer has its best RMSD match (relative to the QM ensemble) with a conformer that has a higher RE.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Table 3. Expanded details of the ability of Auto3D, CREST, Balloon, and ETKDG to match the lowest energy QM structure (global minimum) of Carnosine.

Conformation No.[a]	Auto3D		CREST		Balloon		ETKDG	
	RMSD[b]	Rel_E (rank)	RMSD	Rel_E (rank)	RMSD	Rel_E (rank)	RMSD	Rel_E (rank)
Conf_1	0.07	0.00 (1)	0.31	5.72 (7)	0.57	3.70 (6)	0.49	0.00 (1)
Conf_2	0.19	28.84 (10)	0.33	3.30 (5)	0.72	3.56 (5)	0.73	6.82 (3)
Conf_3	1.10	26.52 (7)	0.57	0.08 (2)	0.79	1.65 (3)	1.36	9.75 (6)
Conf_4	1.54	17.34 (5)	0.61	0.00 (1)	0.90	2.22 (4)	1.54	17.76 (9)
Conf_5	1.57	31.99 (13)	1.10	2.70 (4)	0.90	4.92 (7)	1.87	12.47 (8)
Conf_6	1.59	17.10 (4)	1.49	5.21 (6)	1.07	1.53 (2)	1.90	6.48 (2)
Conf_7	1.67	28.26 (9)	1.51	2.25 (3)	1.33	0.00 (1)	1.92	12.34 (7)
Conf_8	1.70	27.89 (8)			1.37	5.77 (8)	1.96	7.88 (4)
Conf_9	1.73	15.90 (2)			1.45	6.36 (10)	2.01	9.40 (5)
Conf_10	1.83	18.28 (6)			1.88	5.90 (9)		
Conf_11	2.00	16.75 (3)						
Conf_12	2.14	29.13 (11)						
Conf_13	2.74	29.26 (12)						

a) This column is the conformation number for the four conformational search engines. The QM ensemble only contained 12 members (see Table 1).

b) The RMSD column is the RMSD of the generated conformers against the lowest energy QM conformation in ascending order.

c) The Rel E column (in kcal/mol) shows the relative energy of the conformation in the conformational search ensemble and the value in parenthesis shows the rank order of the relative energy within the ensemble. Auto3D and ETKDG were successful (0.00 (1)) with their lowest energy structure having the lowest RMSD to the QM lowest energy structure, while CREST and Balloon represent failures (5.72 (7) and 3.70 (6), respectively).

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Table 4. Predicted CCS and error compared to experiment for the standard QM results, Auto3D, CREST, Balloon, and ETKDG, respectively.[a]

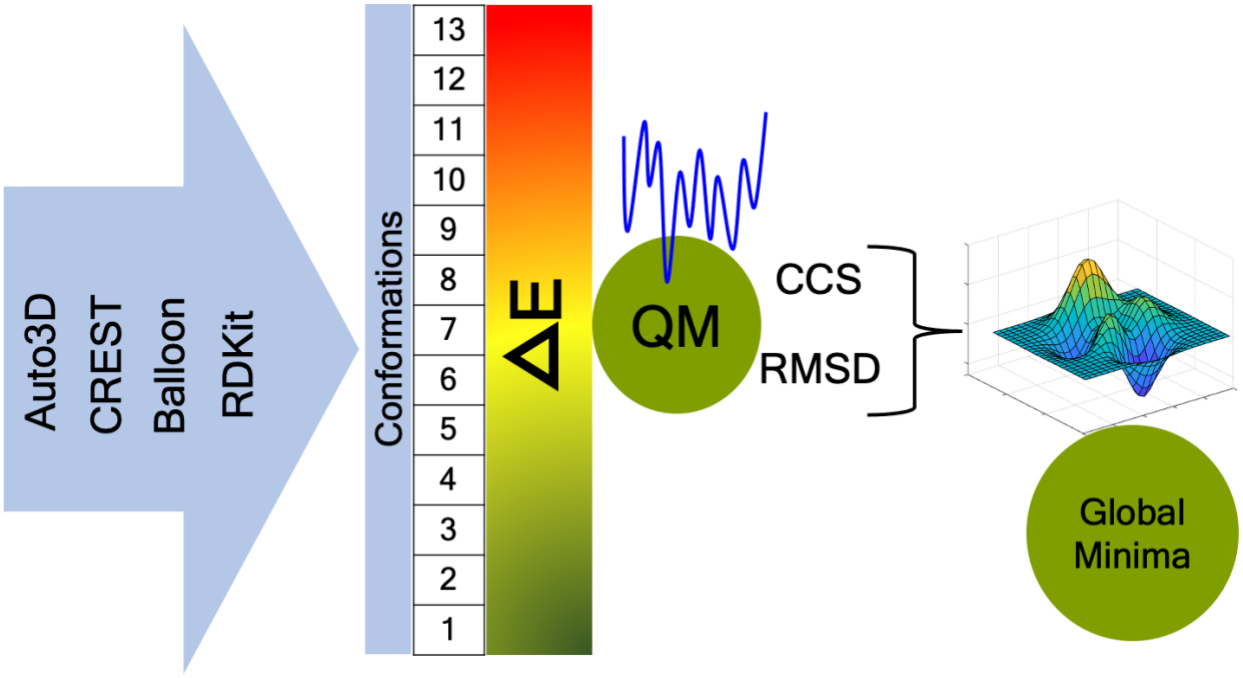
Number	Metabolites	QM (% error)[b]	Auto3D (% error)	CREST (% error)	Balloon (% error)	ETKDG (% error)
1	Carnosine	**150.21 (0.11)**	159.60 (5.99)	**149.52 (0.35)**	161.45 (7.06)	166.49 (9.87)
2	O-succinyl-L-homoserine	**145.45 (0.39)**	**147.86 (2.00)**	**141.03 (2.47)**	**147.41 (1.70)**	138.52 (4.61)
3	L-tyrosine	148.53 (4.14)	153.78 (7.40)	147.12 (3.21)	153.66 (7.33)	150.41 (5.33)
4	L-mimosine	**145.36 (1.43)**	**146.02 (1.86)**	**142.49 (0.57)**	**147.63 (2.93)**	**147.52 (2.86)**
5	citramalic acid	**121.58 (0.24)**	116.51 (4.11)	**118.38 (2.47)**	**118.38 (1.98)**	**120.93 (0.31)**
6	N-methyl-L-glutamate	**129.30 (1.98)**	**133.98 (1.59)**	**128.04 (2.98)**	**135.53 (2.72)**	**135.93 (3.00)**
7	L-ornithine	**127.39 (0.95)**	119.75 (7.39)	123.48 (4.15)	120.18 (7.01)	133.07 (3.36)
8	Abscisic Acid	**162.86 (0.05)**	**166.06 (1.96)**	**160.58 (1.38)**	174.51 (6.71)	173.12 (5.96)
9	L-tryptophan	159.69 (6.06)	161.97 (10.69)	159.62 (9.38)	161.56 (10.47)	167.83 (13.81)
10	L-asparagine	**128.93 (0.20)**	**124.90 (3.00)**	124.20 (3.58)	**126.52 (1.68)**	**125.08 (2.85)**
11	L-anserine	159.74 (3.70)	**155.98 (1.37)**	**152.90 (0.62)**	166.13 (7.39)	171.23 (10.15)
12	Kynurenine	**146.94 (0.47)**	156.59 (5.71)	160.65 (8.09)	159.56 (7.46)	157.45 (6.22)
13	Serotonin	**131.36 (0.38)**	158.29 (6.47)	158.89 (6.82)	159.79 (7.35)	168.18 (11.97)
14	N,N-Dimethylglycine	118.19 (6.23)	113.91 (10.22)	115.02 (9.15)	114.34 (9.80)	116.42 (7.84)
15	L-citrulline	141.86 (4.59)	154.66 (12.49)	**135.55 (0.15)**	161.73 (16.31)	159.52 (15.15)
16	Glutamine	**129.92 (0.58)**	126.11 (3.64)	125.95 (3.77)	124.64 (4.86)	**132.75 (1.54)**
17	L-2-Aminoadipic Acid	**129.56 (1.55)**	137.00 (3.97)	**128.24 (2.58)**	138.36 (4.92)	142.40 (7.62)
18	Guanidinoacetic Acid	**130.21 (2.43)**	120.17 (5.72)	120.23 (5.67)	123.22 (3.11)	122.31 (3.88)
19	Nicotinic Acid	132.17 (3.55)	**127.16 (0.27)**	**126.49 (0.80)**	132.29 (3.62)	**128.64 (0.89)**
20	Quinolinic Acid	142.14 (5.00)	**138.84 (2.73)**	143.91 (6.16)	140.96 (4.19)	140.84 (4.11)

a) If the error is within 3%, the entry is bold, indicating accurate CCS prediction. If the error is greater than 3%, it is not bolded, indicating a poorer prediction of CCS value.

b) The left handed number is the computed CCS value based on a Boltzmann weighting of the ensemble, while the second number is the percent error.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

TOC graphic



https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Reference:

(1) Tsagkaris, C.; Corriero, A. C.; Rayan, R. A.; Moysidis, D. V.; Papazoglou, A. S.; Alexiou, A. Success Stories in Computer-Aided Drug Design. In *Computational Approaches in Drug Discovery, Development and Systems Pharmacology*; Elsevier, 2023; 237–253. https://doi.org/10.1016/B978-0-323-99137-7.00001-0.

(2) Ragno, R.; Esposito, V.; Di Mario, M.; Masiello, S.; Viscovo, M.; Cramer, R. D. Teaching and Learning Computational Drug Design: Student Investigations of 3D Quantitative Structure-Activity Relationships through Web Applications. *J Chem Educ* **2020**, *97*, 1922–1930. https://doi.org/10.1021/ACS.JCHEMED.0C00117/SUPPL_FILE/ED0C00117_SI_001.PDF.

(3) Goudzal, A.; El Aissouq, A.; El Hamdani, H.; Hadaji, E. G.; Ouammou, A.; Bouachrine, M. 3D-QSAR Modeling and Molecular Docking Studies on a Series of 2, 4, 5-Trisubstituted Imidazole Derivatives as CK2 Inhibitors. *https://doi.org/10.1080/07391102.2021.2014360* **2022**, *41*, 234–248. https://doi.org/10.1080/07391102.2021.2014360.

(4) Khaldan, A.; Bouamrane, S.; El-Mernissi, R.; Maghat, H.; Ajana, M. A.; Sbai, A.; Bouachrine, M.; Lakhlifi, T. 3D-QSAR Modeling, Molecular Docking and ADMET Properties of Benzothiazole Derivatives as α-Glucosidase Inhibitors. *Mater Today Proc* **2021**, *45*, 7643–7652. https://doi.org/10.1016/J.MATPR.2021.03.114.

(5) Guterres, H.; Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J Chem Inf Model* **2020**, *60*, 2189–2198. https://doi.org/10.1021/ACS.JCIM.0C00057/ASSET/IMAGES/LARGE/CI0C00057_0005.JPEG.

(6) Bao, J.; He, X.; Zhang, J. Z. H. DeepBSP-a Machine Learning Method for Accurate Prediction of Protein-Ligand Docking Structures. *J Chem Inf Model* **2021**, *61* (5), 2231–2240. https://doi.org/10.1021/ACS.JCIM.1C00334/ASSET/IMAGES/LARGE/CI1C00334_0009.JPEG.

(7) Miteva, M. A.; Guyon, F.; Tufféry, P. Frog2: Efficient 3D Conformation Ensemble Generator for Small Compounds. *Nucleic Acids Res* **2010**, *38*, 622–627. https://doi.org/10.1093/NAR/GKQ325.

(8) Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. EnzyDock: Protein–Ligand Docking of Multiple Reactive States along a Reaction Coordinate in Enzymes. *J Chem Theory Comput* **2019**, *15*, 5116–5134. https://doi.org/10.1021/acs.jctc.9b00366.

(9) Das, S.; Edison, A. S.; Merz, K. M. Metabolite Structure Assignment Using in Silico NMR Techniques. *Anal Chem* **2020**, *92*, 10412–10419. https://doi.org/10.1021/ACS.ANALCHEM.0C00768/ASSET/IMAGES/LARGE/AC0C00768_0005.JPEG.

(10) Das, S.; Tanemura, K. A.; Dinpazhoh, L.; Keng, M.; Schumm, C.; Leahy, L.; Asef, C. K.; Rainey, M.; Edison, A. S.; Fernández, F. M.; Merz, K. M. In Silico Collision Cross Section Calculations to Aid

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Metabolite Annotation. *J Am Soc Mass Spectrom* **2022**, *33*, 750–759. https://doi.org/10.1021/JASMS.1C00315/ASSET/IMAGES/LARGE/JS1C00315_0004.JPEG.

(11) Borges, R. M.; Colby, S. M.; Das, S.; Edison, A. S.; Fiehn, O.; Kind, T.; Lee, J.; Merrill, A. T.; Merz Jr., K. M.; Metz, T. O.; Nunez, J. R.; Tantillo, D. J.; Wang, L. P.; Wang, S.; Renslow, R. S. Quantum Chemistry Calculations for Metabolomics. *Chem Rev* **2021**, *121*, 5633–5670. https://doi.org/10.1021/acs.chemrev.0c00901.

(12) Puranen, J. S.; Vainio, M. J.; Johnson, M. S. Accurate Conformation-Dependent Molecular Electrostatic Potentials for High-Throughput in Silico Drug Discovery. *J Comput Chem* **2010**, *31*, 1722–1732. https://doi.org/10.1002/JCC.21460.

(13) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J Chem Inf Model* **2007**, *47* (6), 2462–2474. https://doi.org/10.1021/ci6005646.

(14) *RDKit*. https://doi.org/https://www.rdkit.org/.

(15) Landrum, G.; Tosco, P.; Kelley, B.; Ric; sriniker; Cosgrove, D.; gedeck; Vianello, R.; NadineSchneider; Kawashima, E.; N, D.; Jones, G.; Dalke, A.; Cole, B.; Swain, M.; Turk, S.; AlexanderSavelyev; Vaucher, A.; Wójcikowski, M.; Take, I.; Probst, D.; Ujihara, K.; Scalfani, V. F.; godin, guillaume; Pahl, A.; Berenger, F.; JLVarjo; Walker, R.; jasondbiggs; strets123. Rdkit/Rdkit: Release. **2023**. https://doi.org/10.5281/ZENODO.7880616.

(16) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J Chem Inf Model* **2015**, *55*, 2562–2574. https://doi.org/10.1021/ACS.JCIM.5B00654.

(17) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule-Ligand Interactions. *J Mol Biol* **1982**, *161*, 269–288. https://doi.org/10.1016/0022-2836(82)90153-X.

(18) Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tufféry, P. Frog: A FRee Online DruG 3D Conformation Generator. *Nucleic Acids Res* **2007**, *35*, 568–572. https://doi.org/10.1093/NAR/GKM289.

(19) Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with Macromodel. *J Chem Inf Model* **2014**, *54*, 2680–2696. https://doi.org/10.1021/CI5001696/SUPPL_FILE/CI5001696_SI_003.TXT.

(20) Kolossváry, I.; Guida, W. C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *J Am Chem Soc* **1996**, *118*, 5011–5019. https://doi.org/10.1021/JA952478M/ASSET/IMAGES/LARGE/JA952478MF00004.JPEG.

(21) 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

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Databank and Cambridge Structural Database. *J Chem Inf Model* **2010**, *50*, 572–584. https://doi.org/10.1021/CI100031X/SUPPL_FILE/CI100031X_SI_004.TXT.

(22) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the Performance of OMEGA with Respect to Retrieving Bioactive Conformations. *J Mol Graph Model* **2003**, *21*, 449–462. https://doi.org/10.1016/S1093-3263(02)00204-8.

(23) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Physical Chemistry Chemical Physics* **2020**, *22*, 7169–7192. https://doi.org/10.1039/C9CP06869D.

(24) Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S. Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods. *J Chem Theory Comput* **2022**, *18*, 3189. https://doi.org/10.1021/ACS.JCTC.2C00239/SUPPL_FILE/CT2C00239_SI_002.ZIP.

(25) Plett, C.; Grimme, S. Automated and Efficient Generation of General Molecular Aggregate Structures. *Angewandte Chemie* **2023**, *135*, 202214477 - 202217783. https://doi.org/10.1002/ange.202214477.

(26) Liu, Z.; Zubatiuk, T.; Roitberg, A.; Isayev, O. Auto3D: Automatic Generation of the Low-Energy 3D Structures with ANI Neural Network Potentials. *J Chem Inf Model* **2022**, *62*, 5373–5382. https://doi.org/10.1021/ACS.JCIM.2C00817/SUPPL_FILE/CI2C00817_SI_002.ZIP.

(27) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Zhou, Z.; Han, L.; Karapetyan, K.; Dracheva, S.; Shoemaker, B. A.; Bolton, E.; Gindulyte, A.; Bryant, S. H. PubChem's BioAssay Database. *Nucleic Acids Res* **2012**, *40*, 400–412. https://doi.org/10.1093/nar/gkr1132.

(28) Politi, R.; Rusyn, I.; Tropsha, A. Prediction of Binding Affinity and Efficacy of Thyroid Hormone Receptor Ligands Using QSAR and Structure-Based Modeling Methods. *Toxicol Appl Pharmacol* **2014**, *280*, 177–189. https://doi.org/10.1016/j.taap.2014.07.009.

(29) Karwath, A.; De Raedt, L. SMIREP: Predicting Chemical Activity from SMILES. *J Chem Inf Model* **2006**, *46*, 2432–2444. https://doi.org/10.1021/ci060159g.

(30) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J Cheminform* **2011**, *3* (10), 1–14. https://doi.org/10.1186/1758-2946-3-33/TABLES/2.

(31) Melville, J. L.; Hirst, J. D. TMACC: Interpretable Correlation Descriptors for Quantitative Structure−Activity Relationships. *J Chem Inf Model* **2007**, *47*, 626–634. https://doi.org/10.1021/ci6004178.

(32) Gakh, A. A.; Burnett, M. N.; Trepalin, S. V.; Yarkov, A. V. Modular Chemical Descriptor Language (MCDL): Stereochemical Modules. *J Cheminf* **2011**, *3*, 5 – 14. https://doi.org/10.1186/1758-2946-3-5.

(33) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. Confab - Systematic Generation of Diverse Low-Energy Conformers. *J Cheminf* **2011**, *3*, 8 - 16. https://doi.org/10.1186/1758-2946-3-8.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(34) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J Chem Inf Comput Sci* **1988**, *28* (1), 31–36. https://doi.org/10.1021/ci00057a005.

(35) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: A Python Wrapper for the OpenBabel Cheminformatics Toolkit. *Chem Cent J* **2008**, *2*, 5. https://doi.org/10.1186/1752-153x-2-5.

(36) Douguet, D. Ligand-Based Approaches in Virtual Screening. *Current Computer Aided-Drug Design* **2008**, *4*, 180–190. https://doi.org/10.2174/157340908785747456.

(37) Miteva, M. A.; Violas, S.; Montes, M.; Gomez, D.; Tuffery, P.; Villoutreix, B. O. FAF-Drugs: Free ADME/Tox Filtering of Compound Collections. *Nucleic Acids Res* **2006**, *34*, 738 - 744. https://doi.org/10.1093/NAR/GKL065.

(38) Bender, A.; Jenkins, J. L.; Scheiber, J.; Sukuru, S. C. K.; Glick, M.; Davies, J. W. How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *J Chem Inf Model* **2009**, *49*, 108–119. https://doi.org/10.1021/CI800249S.

(39) Sperandio, O.; Souaille, M.; Delfaud, F.; Miteva, M. A.; Villoutreix, B. O. MED-3DMC: A New Tool to Generate 3D Conformation Ensembles of Small Molecules with a Monte Carlo Sampling of the Conformational Space. *Eur J Med Chem* **2009**, *44*, 1405–1409. https://doi.org/10.1016/J.EJMECH.2008.09.052.

(40) Shoichet, B. K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862–865. https://doi.org/10.1038/NATURE03197.

(41) Clark, D. E. What Has Virtual Screening Ever Done for Drug Discovery? *Expert Opin Drug Discov* **2008**, *3*, 841–851. https://doi.org/10.1517/17460441.3.8.841.

(42) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science (1979)* **1992**, *257*, 1078–1082. https://doi.org/10.1126/SCIENCE.257.5073.1078.

(43) Yang, Y.; Hsieh, C. Y.; Kang, Y.; Hou, T.; Liu, H.; Yao, X. Deep Generation Model Guided by the Docking Score for Active Molecular Design. *J Chem Inf Model* **2023**, *63*, 2983–2991. https://doi.org/10.1021/ACS.JCIM.3C00572.

(44) Kamal, I. M.; Chakrabarti, S. MetaDOCK: A Combinatorial Molecular Docking Approach. *ACS Omega* **2023**, *8*, 5850–5860. https://doi.org/10.1021/ACSOMEGA.2C07619.

(45) Hsu, D. J.; Davidson, R. B.; Sedova, A.; Glaser, J. TinyIFD: A High-Throughput Binding Pose Refinement Workflow Through Induced-Fit Ligand Docking. *J Chem Inf Model* **2023**, *63*, 3438–3447. https://doi.org/10.1021/ACS.JCIM.2C01530.

(46) Alkhodier, R. A.; Mishra, S. K.; Doerksen, R. J.; Colby, D. A. Comparison of Conformational Analyses of Naturally Occurring Flavonoid-O-Glycosides with Unnatural Flavonoid-CF2-Glycosides Using Molecular Modeling. *J Chem Inf Model* **2023**, *63*, 375–386. https://doi.org/10.1021/ACS.JCIM.2C01147.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(47) Chia, S.; Faidon Brotzakis, Z.; Horne, R. I.; Possenti, A.; Mannini, B.; Cataldi, R.; Nowinska, M.; Staats, R.; Linse, S.; Knowles, T. P. J.; Habchi, J.; Vendruscolo, M. Structure-Based Discovery of Small-Molecule Inhibitors of the Autocatalytic Proliferation of α-Synuclein Aggregates. *Mol Pharm* **2023**, *20*, 183–193. https://doi.org/10.1021/ACS.MOLPHARMACEUT.2C00548.

(48) Yu, Y.; Cai, C.; Wang, J.; Bo, Z.; Zhu, Z.; Zheng, H. Uni-Dock: GPU-Accelerated Docking Enables Ultralarge Virtual Screening. *J Chem Theory Comput* **2023,** 19, 3336 – 3345. https://doi.org/10.1021/acs.jctc.2c01145.

(49) Perez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience. *Chem Rev* **2022**, *122*, 15914–15970. https://doi.org/10.1021/ACS.CHEMREV.2C00060.

(50) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem* **2004**, *47*, 1739–1749. https://doi.org/10.1021/JM0306430/SUPPL_FILE/JM0306430_S.PDF.

(51) Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C. Macrocycle Conformational Sampling with Macromodel. *J Chem Inf Model* **2014**, *54*, 2680–2696. https://doi.org/10.1021/CI5001696.

(52) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem Rev* **1997**, *97*, 1359–1472. https://doi.org/10.1021/CR960370Z.

(53) Shim, J.; MacKerell, A. D. Computational Ligand-Based Rational Design: Role of Conformational Sampling and Force Fields in Model Development. *Medchemcomm* **2011**, *2*, 356–370. https://doi.org/10.1039/C1MD00044F.

(54) 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. *J Chem Inf Model* **2007**, *47*, 1923–1932. https://doi.org/10.1021/CI700136X.

(55) Hawkins, P. C. D. Conformation Generation: The State of the Art. *J Chem Inf Model* **2017**, *57*, 1747–1756. https://doi.org/10.1021/ACS.JCIM.7B00221.

(56) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J Chem Inf Model* **2005**, *45*, 160–169. https://doi.org/10.1021/CI049885E.

(57) Smellie, A.; Stanton, R.; Henne, R.; Teig, S. Conformational Analysis by Intersection: CONAN. *J Comput Chem* **2003**, *24*, 10–20. https://doi.org/10.1002/JCC.10175.

(58) Schwab, C. H. Conformations and 3D Pharmacophore Searching. *Drug Discov Today Technol* **2010**, *7*. https://doi.org/10.1016/J.DDTEC.2010.10.003.

(59) 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

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Databank and Cambridge Structural Database. *J Chem Inf Model* **2010**, *50*, 572–584. https://doi.org/10.1021/CI100031X.

(60) Hawkins, P. C. D.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J Chem Inf Model* **2012**, *52*, 2919–2936. https://doi.org/10.1021/ci300314k.

(61) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J Chem Inf Model* **2010**, *50*, 534–546. https://doi.org/10.1021/CI100015J.

(62) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J Med Chem* **2007**, *50*, 74–82. https://doi.org/10.1021/JM0603365.

(63) Poli, G.; Seidel, T.; Langer, T. Conformational Sampling of Small Molecules with ICon: Performance Assessment in Comparison with OMEGA. *Front Chem* **2018**, *6*, 349617. https://doi.org/10.3389/FCHEM.2018.00229/BIBTEX.

(64) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Physical Chemistry Chemical Physics* **2020**, *22*, 7169–7192. https://doi.org/10.1039/C9CP06869D.

(65) Sheong, F. K.; Zhang, J. X.; Lin, Z. Localized Bonding Model for Coordination and Cluster Compounds. *Coord Chem Rev* **2017**, *345*, 42–53. https://doi.org/10.1016/j.ccr.2016.10.012.

(66) Mahmudov, K. T.; Kopylovich, M. N.; Guedes da Silva, M. F. C.; Pombeiro, A. J. L. Non-Covalent Interactions in the Synthesis of Coordination Compounds: Recent Advances. *Coord Chem Rev* **2017**, *345*, 54–72. https://doi.org/10.1016/J.CCR.2016.09.002.

(67) Harrison, J. A.; Schall, J. D.; Maskey, S.; Mikulski, P. T.; Knippenberg, M. T.; Morrow, B. H. Review of Force Fields and Intermolecular Potentials Used in Atomistic Computational Materials Research. *Appl Phys Rev* **2018**, *5*, 031104 – 031127. https://doi.org/10.1063/1.5020808.

(68) Bursch, M.; Mewes, J. M.; Hansen, A.; Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry. *Angewandte Chemie - International Edition* **2022**, *61* (42), 2022057 - 2022084. https://doi.org/10.1002/ANIE.202205735.

(69) Mahmudov, K. T.; Kopylovich, M. N.; Guedes da Silva, M. F. C.; Pombeiro, A. J. L. Non-Covalent Interactions in the Synthesis of Coordination Compounds: Recent Advances. *Coord Chem Rev* **2017**, *345*, 54–72. https://doi.org/10.1016/J.CCR.2016.09.002.

(70) Plett, C.; Grimme, S. Automated and Efficient Generation of General Molecular Aggregate Structures. *Angewandte Chemie* **2023**, *135*, e202214477. https://doi.org/10.1002/ANGE.202214477.

(71) Young, D.; Martin, T.; Venkatapathy, R.; Harten, P. Are the Chemical Structures in Your QSAR Correct? *QSAR Comb Sci* **2008**, *27*, 1337–1345. https://doi.org/10.1002/qsar.200810084.

(72) Poli, G.; Seidel, T.; Langer, T. Conformational Sampling of Small Molecules With ICon: Performance Assessment in Comparison With OMEGA. *Front Chem* **2018**, *6*. 1 – 18. https://doi.org/10.3389/fchem.2018.00229.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(73) 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. https://doi.org/10.1021/ACS.JCTC.8B01176.

(74) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat Commun* **2019**, *10*, 1 -8. https://doi.org/10.1038/S41467-019-10827-4.

(75) Gupta, A.; Zhou, H. X. Machine Learning-Enabled Pipeline for Large-Scale Virtual Drug Screening. *J Chem Inf Model* **2021**, *61*, 4236–4244. https://doi.org/10.1021/ACS.JCIM.1C00710.

(76) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. *J Chem Inf Model* **2020**, *60*, 3408–3415. https://doi.org/10.1021/ACS.JCIM.0C00451.

(77) Ji, Z.; Shi, R.; Lu, J.; Li, F.; Yang, Y. ReLMole: Molecular Representation Learning Based on Two-Level Graph Similarities. *J Chem Inf Model* **2022**, *62*, 5361–5372. https://doi.org/10.1021/ACS.JCIM.2C00798.

(78) Liu, Z.; Zubatiuk, T.; Roitberg, A.; Isayev, O. Auto3D: Automatic Generation of the Low-Energy 3D Structures with ANI Neural Network Potentials. *J Chem Inf Model* **2022**, *62*, 5373–5382. https://doi.org/10.1021/ACS.JCIM.2C00817/SUPPL_FILE/CI2C00817_SI_002.ZIP.

(79) Pan, X.; Zhao, F.; Zhang, Y.; Wang, X.; Xiao, X.; Zhang, J. Z. H.; Ji, C. MolTaut: A Tool for the Rapid Generation of Favorable Tautomer in Aqueous Solution. *J Chem Inf Model* **2023,** *63*, 1833–1840. https://doi.org/10.1021/ACS.JCIM.2C01393.

(80) Mendenhall, J.; Brown, B. P.; Kothiwale, S.; Meiler, J. BCL::Conf: Improved Open-Source Knowledge-Based Conformation Sampling Using the Crystallography Open Database. *J Chem Inf Model* **2021**, *61*, 189–201. https://doi.org/10.1021/acs.jcim.0c01140.

(81) 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*, 2044–2058. https://doi.org/10.1021/ACS.JCIM.0C00025/SUPPL_FILE/CI0C00025_SI_004.ZIP.

(82) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J Chem Inf Model* **2019**, *59*, 731–742. https://doi.org/10.1021/acs.jcim.8b00704.

(83) Cole, J. C.; Korb, O.; McCabe, P.; Read, M. G.; Taylor, R. Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *J Chem Inf Model* **2018**, *58*, 615–629. https://doi.org/10.1021/ACS.JCIM.7B00697/SUPPL_FILE/CI7B00697_SI_003.ZIP.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(84) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Computer Methodology* **1990**, *3*, 537–547. https://doi.org/10.1016/0898-5529(90)90156-3.

(85) *Chemical Computing Group (CCG) | Computer-Aided Molecular Design*. https://www.chemcomp.com/ (accessed 2023-07-17).

(86) Seidel, T.; Permann, C.; Wieder, O.; Kohlbacher, S. M.; Langer, T. High-Quality Conformer Generation with CDPKit/CONFORT: Algorithm and Performance Assessment. **2023,** *63*, 5549–5570. https://doi.org/10.21203/RS.3.RS-1597257/V1.

(87) Dressler, F.; Dietrich, I.; German, R.; Krüger, B. A Rule-Based System for Programming Self-Organized Sensor and Actor Networks. *Computer Networks* **2009**, *53*, 1737–1750. https://doi.org/10.1016/J.COMNET.2008.09.007.

(88) Slowik, A.; Kwasnicka, H. Evolutionary Algorithms and Their Applications to Engineering Problems. *Neural Comput Appl* **2020**, *32*, 12363–12379. https://doi.org/10.1007/S00521-020-04832-8/TABLES/4.

(89) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting Conformational Variation. *J Comput Chem* **1995**, *16*, 171–187. https://doi.org/10.1002/JCC.540160205.

(90) Smellie, A.; Kahn, S. D.; Teig, S. L. Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage. *J Chem Inf Comput Sci* **1995**, *35*, 285–294. https://doi.org/10.1021/CI00024A018/ASSET/CI00024A018.FP.PNG_V03.

(91) Sperandio, O.; Souaille, M.; Delfaud, F.; Miteva, M. A.; Villoutreix, B. O. MED-3DMC: A New Tool to Generate 3D Conformation Ensembles of Small Molecules with a Monte Carlo Sampling of the Conformational Space. *Eur J Med Chem* **2009**, *44*, 1405–1409. https://doi.org/10.1016/J.EJMECH.2008.09.052.

(92) Sauton, N.; Lagorce, D.; Villoutreix, B. O.; Miteva, M. A. MS-DOCK: Accurate Multiple Conformation Generator and Rigid Docking Protocol for Multi-Step Virtual Ligand Screening. *BMC Bioinformatics* **2008**, *9*, 1–12. https://doi.org/10.1186/1471-2105-9-184/FIGURES/3.

(93) Schärfer, C.; Schulz-Gasch, T.; Hert, J.; Heinzerling, L.; Schulz, B.; Inhester, T.; Stahl, M.; Rarey, M. CONFECT: Conformations from an Expert Collection of Torsion Patterns. *ChemMedChem* **2013**, *8*, 1690–1700. https://doi.org/10.1002/CMDC.201300242.

(94) Coutsias, E. A.; Lexa, K. W.; Wester, M. J.; Pollock, S. N.; Jacobson, M. P. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics. *J Chem Theory Comput* **2016**, *12*, 4674–4687. https://doi.org/10.1021/ACS.JCTC.6B00250/SUPPL_FILE/CT6B00250_SI_002.PDF.

(95) Cleves, A. E.; Jain, A. N. ForceGen 3D Structure and Conformer Generation: From Small Lead-like Molecules to Macrocyclic Drugs. *J Comput Aided Mol Des* **2017**, *31*, 419–439. https://doi.org/10.1007/S10822-017-0015-8/FIGURES/13.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(96) Griewel, A.; Kayser, O.; Schlosser, J.; Rarey, M. Conformational Sampling for Large-Scale Virtual Screening: Accuracy versus Ensemble Size. *J Chem Inf Model* **2009**, *49*, 2303–2311. https://doi.org/10.1021/CI9002415/ASSET/IMAGES/LARGE/CI-2009-002415_0009.JPEG.

(97) Friedrich, N. O.; De Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J Chem Inf Model* **2017**, *57*, 2719–2728. https://doi.org/10.1021/ACS.JCIM.7B00505/SUPPL_FILE/CI7B00505_SI_003.TXT.

(98) Friedrich, N. O.; Meyder, A.; De Bruyn Kops, C.; Sommer, K.; Flachsenberg, F.; Rarey, M.; Kirchmair, J. High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators. *J Chem Inf Model* **2017**, *57*, 529–539. https://doi.org/10.1021/ACS.JCIM.6B00613/SUPPL_FILE/CI6B00613_SI_003.TXT.

(99) Beutler, T. C.; Dill, K. A. A Fast Conformational Search Strategy for Finding Low Energy Structures of Model Proteins. *Protein Science* **1996**, *5*, 2037–2043. https://doi.org/10.1002/PRO.5560051010.

(100) Beutler, T. C.; Dill, K. A. A Fast Conformational Search Strategy for Finding Low Energy Structures of Model Proteins. *Protein Science* **1996**, *5*, 2037–2043. https://doi.org/10.1002/PRO.5560051010.

(101) Garman, E. F.; Schneider, T. R. Macromolecular Cryocrystallography. *J Appl Crystallogr* **1997**, *30*, 211–237. https://doi.org/10.1107/S0021889897002677.

(102) Miller, R. J. D. Femtosecond Crystallography with Ultrabright Electrons and X-Rays: Capturing Chemistry in Action. *Science (1979)* **2014**, *343*, 1108–1116. https://doi.org/10.1126/SCIENCE.1248488.

(103) Garman, E. F. Developments in X-Ray Crystallographic Structure Determination of Biological Macromolecules. *Science (1979)* **2014**, *343*, 1102–1108. https://doi.org/10.1126/SCIENCE.1247829.

(104) Amanullah, A.; Naheed, S. Structural Bioinformatics: Computational Software and Databases for the Evaluation of Protein Structure. *RADS Journal of Biological Research & Applied Sciences* **2018**, *9*, 94–101. https://doi.org/10.37962/JBAS.V9I2.124.

(105) Grauso, L.; Teta, R.; Esposito, G.; Menna, M.; Mangoni, A. Computational Prediction of Chiroptical Properties in Structure Elucidation of Natural Products. *Nat Prod Rep* **2019**, *36*, 1005–1030. https://doi.org/10.1039/C9NP00018F.

(106) Mándi, A.; Kurtán, T. Applications of OR/ECD/VCD to the Structure Elucidation of Natural Products. *Nat Prod Rep* **2019**, *36*, 889–918. https://doi.org/10.1039/C9NP00002J.

(107) Burns, D. C.; Mazzola, E. P.; Reynolds, W. F. The Role of Computer-Assisted Structure Elucidation (CASE) Programs in the Structure Elucidation of Complex Natural Products. *Nat Prod Rep* **2019**, *36*, 919–933. https://doi.org/10.1039/C9NP00007K.

(108) Linington, R. G.; Kubanek, J.; Luesch, H. New Methods for Isolation and Structure Determination of Natural Products. *Nat Prod Rep* **2019**, *36*, 942–943. https://doi.org/10.1039/C9NP90023C.

(109) Wilkins, S. W. Celebrating 100 Years of X-Ray Crystallography. *Acta Crystallogr A* **2013**, *69*, 1–4. https://doi.org/10.1107/S0108767312048490/ME0486SUP4.PDF.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(110) *RCSB PDB: Homepage*. https://www.rcsb.org/ (accessed 2023-07-17).

(111) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J Med Chem* **2004**, *47*, 2499–2510. https://doi.org/10.1021/jm030563w.

(112) Zivanovic, S.; Bayarri, G.; Colizzi, F.; Moreno, D.; Gelpí, J. L.; Soliva, R.; Hospital, A.; Orozco, M. Bioactive Conformational Ensemble Server and Database. A Public Framework to Speed up in Silico Drug Discovery. *J Chem Theory Comput* **2020**, *16*, 6586–6597. https://doi.org/10.1021/ACS.JCTC.0C00305.

(113) Zivanovic, S.; Colizzi, F.; Moreno, D.; Hospital, A.; Soliva, R.; Orozco, M. Exploring the Conformational Landscape of Bioactive Small Molecules. *J Chem Theory Comput* **2020**, *16*, 6575–6585. https://doi.org/10.1021/ACS.JCTC.0C00304/ASSET/IMAGES/LARGE/CT0C00304_0010.JPEG.

(114) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J Med Chem* **2004**, *47*, 2499–2510. https://doi.org/10.1021/JM030563W/ASSET/IMAGES/LARGE/JM030563WF00004.JPEG.

(115) Pearlman, D. A.; Charifson, P. S. Improved Scoring of Ligand-Protein Interactions Using OWFEG Free Energy Grids. *J Med Chem* **2001**, *44*, 502–511. https://doi.org/10.1021/JM000375V/ASSET/IMAGES/MEDIUM/JM000375VE00010.GIF.

(116) Vieth, M.; Hirst, J. D.; Brooks, C. L. Do Active Site Conformations of Small Ligands Correspond to Low Free-Energy Solution Structures? *J Comput Aided Mol Des* **1998**, *12*, 563–572. https://doi.org/10.1023/A:1008055202136/METRICS.

(117) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding. *J Med Chem* **2004**, *47*, 2499–2510. https://doi.org/10.1021/JM030563W/ASSET/IMAGES/LARGE/JM030563WF00004.JPEG.

(118) Pearlman, D. A.; Charifson, P. S. Improved Scoring of Ligand-Protein Interactions Using OWFEG Free Energy Grids. *J Med Chem* **2001**, *44*, 502–511. https://doi.org/10.1021/JM000375V/ASSET/IMAGES/MEDIUM/JM000375VE00010.GIF.

(119) Lyne, P. D. Structure-Based Virtual Screening: An Overview. *Drug Discov Today* **2002**, *7*, 1047–1055. https://doi.org/10.1016/S1359-6446(02)02483-2.

(120) Butler, K. T.; Luque, F. J.; Barril, X. Toward Accurate Relative Energy Predictions of the Bioactive Conformation of Drugs. *J Comput Chem* **2009**, *30*, 601–610. https://doi.org/10.1002/jcc.21087.

(121) Yu, Z.; Li, P.; Merz, K. M. Using Ligand-Induced Protein Chemical Shift Perturbations To Determine Protein-Ligand Structures. *Biochemistry* **2017**, *56*, 2349–2362. https://doi.org/10.1021/ACS.BIOCHEM.7B00170/ASSET/IMAGES/LARGE/BI-2017-001707_0010.JPEG.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(122) Fu, Z.; Li, X.; Merz, K. M. Conformational Analysis of Free and Bound Retinoic Acid. *J Chem Theory Comput* **2012**, *8*, 1436–1448. https://doi.org/10.1021/CT200813Q/SUPPL_FILE/CT200813Q_SI_001.PDF.

(123) Pan, L. L.; Zheng, Z.; Wang, T.; Merz, K. M. Free Energy-Based Conformational Search Algorithm Using the Movable Type Sampling Method. *J Chem Theory Comput* **2015**, *11*, 5853–5864. https://doi.org/10.1021/ACS.JCTC.5B00930/SUPPL_FILE/CT5B00930_SI_003.TXT.

(124) Fu, Z.; Li, X.; Miao, Y.; Merz, K. M. Conformational Analysis and Parallel QM/MM X-Ray Refinement of Protein Bound Anti-Alzheimer Drug Donepezil. *J Chem Theory Comput* **2013**, *9*, 1686–1693. https://doi.org/10.1021/ct300957x.

(125) Li, X.; He, X.; Wang, B.; Merz, K. Conformational Variability of Benzamidinium-Based Inhibitors. *J Am Chem Soc* **2009**, *131*, 7742–7754. https://doi.org/10.1021/JA9010833/SUPPL_FILE/JA9010833_SI_001.PDF.

(126) Li, P.; Song, L. F.; Merz, K. M. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J Chem Theory Comput* **2015**, *11*, 1645–1657. https://doi.org/10.1021/CT500918T/SUPPL_FILE/CT500918T_SI_003.XLSX.

(127) Zheng, H.; Hou, J.; Zimmerman, M. D.; Wlodawer, A.; Minor, W. The Future of Crystallography in Drug Discovery. *Expert Opin Drug Discov* **2014**, *9*, 125–137. https://doi.org/10.1517/17460441.2014.872623.

(128) Shabalin, I.; Dauter, Z.; Jaskolski, M.; Minor, W.; Wlodawer, A. Crystallography and Chemistry Should Always Go Together: A Cautionary Tale of Protein Complexes with Cisplatin and Carboplatin. *Acta Crystallogr D Biol Crystallogr* **2015**, *71*, 1965–1979. https://doi.org/10.1107/S139900471500629X.

(129) Zheng, H.; Handing, K. B.; Zimmerman, M. D.; Shabalin, I. G.; Almo, S. C.; Minor, W. X-Ray Crystallography over the Past Decade for Novel Drug Discovery – Where Are We Heading Next? *http://dx.doi.org/10.1517/17460441.2015.1061991* **2015**, *10*, 975–989. https://doi.org/10.1517/17460441.2015.1061991.

(130) Cooper, D. R.; Porebski, P. J.; Chruszcz, M.; Minor, W. X-Ray Crystallography: Assessment and Validation of Proteinsmall Molecule Complexes for Drug Discovery. *Expert Opin Drug Discov* **2011**, *6*, 771–782. https://doi.org/10.1517/17460441.2011.585154.

(131) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *urn:issn:2052-5206* **2016**, *72*, 171–179. https://doi.org/10.1107/S2052520616003954.

(132) Eiceman, G. A.; Karpas, Z. *Ion Mobility Spectrometry*; CRC Press., 2005. https://doi.org/10.1201/9781420038972.

(133) Giles, K.; Williams, J. P.; Campuzano, I. Enhancements in Travelling Wave Ion Mobility Resolution. *Rapid Communications in Mass Spectrometry* **2011**, *25*, 1559–1566. https://doi.org/10.1002/rcm.5013.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(134) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal Chem* **2015**, *87*, 1422–1436. https://doi.org/10.1021/ac504720m.

(135) Michelmann, K.; Silveira, J. A.; Ridgeway, M. E.; Park, M. A. Fundamentals of Trapped Ion Mobility Spectrometry. *J Am Soc Mass Spectrom* **2015**, *26*, 14–24. https://doi.org/10.1007/s13361-014-0999-4.

(136) Basanta, M.; Jarvis, R. M.; Xu, Y.; Blackburn, G.; Tal-Singer, R.; Woodcock, A.; Singh, D.; Goodacre, R.; Thomas, C. L.; Fowler, S. J. Non-Invasive Metabolomic Analysis of Breath Using Differential Mobility Spectrometry in Patients with Chronic Obstructive Pulmonary Disease and Healthy Smokers. *Analyst* **2010**, *135*, 315–320. https://doi.org/10.1039/b916374c.

(137) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem Phys Lett* **1996**, *261*, 86–91. https://doi.org/Doi 10.1016/0009-2614(96)00941-4.

(138) Ahmed, A.; Ju Cho, Y.; No, M.; Koh, J.; Tomczyk, N.; Giles, K.; Shin Yoo, J.; Kim, S. Application of the Mason−Schamp Equation and Ion Mobility Mass Spectrometry To Identify Structurally Related Compounds in Crude Oil. *Anal Chem* **2010**, *83*, 77–83. https://doi.org/10.1021/ac101934q.

(139) E. Revercomb, H.; A. Mason, E. Theory of Plasma Chromatography/Gaseous Electrophoresis. Review. *Anal Chem* **2002**, *47*, 970–983. https://doi.org/10.1021/ac60357a043.

(140) Wyttenbach, T.; von Helden, G.; J. Batka, J.; Carlat, D.; T. Bowers, M. Effect of the Long-Range Potential on Ion Mobility Measurements. *J Am Soc Mass Spectrom* **1997**, *8*, 275–282. https://doi.org/10.1016/S1044-0305(96)00236-X.

(141) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *Journal of Physical Chemistry* **1996**, *100*, 16082–16086. https://doi.org/DOI 10.1021/jp961623v.

(142) Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Anal Chem* **2014**, *86*, 3985–3993. https://doi.org/10.1021/AC500405X.

(143) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res* **2016**, *44*, D1202–D1213. https://doi.org/10.1093/NAR/GKV951.

(144) Hong, H.; Xie, Q.; Ge, W.; Qian, F.; Fang, H.; Shi, L.; Su, Z.; Perkins, R.; Tong, W. Mold2, Molecular Descriptors from 2D Structures for Chemoinformatics and Toxicoinformatics. *J Chem Inf Model* **2008**, *48*, 1337–1344. https://doi.org/10.1021/CI800038F.

(145) Soper-Hopper, M. T.; Vandegrift, J.; Baker, E. S.; Fernández, F. M. Metabolite Collision Cross Section Prediction without Energy-Minimized Structures. *Analyst* **2020**, *145*, 5414–5418. https://doi.org/10.1039/D0AN00198H.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(146) Monge, M. E.; Dodds, J. N.; Baker, E. S.; Edison, A. S.; Fernaacutendez, F. M. Challenges in Identifying the Dark Molecules of Life. *Annual Review of Analytical Chemistry* **2019**, *12*, 177–199. https://doi.org/10.1146/ANNUREV-ANCHEM-061318-114959.

(147) Gabelica, V.; Marklund, E. Fundamentals of Ion Mobility Spectrometry. *Curr Opin Chem Biol* **2018**, *42*, 51–59. https://doi.org/10.1016/J.CBPA.2017.10.022.

(148) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem Sci* **2017**, *8*, 3192–3203. https://doi.org/10.1039/C6SC05720A.

(149) Tanemura, K. A.; Das, S.; Merz, K. M. AutoGraph: Autonomous Graph-Based Clustering of Small-Molecule Conformations. *J Chem Inf Model* **2021**, *61*, 1647–1656. https://doi.org/10.1021/ACS.JCIM.0C01492/ASSET/IMAGES/LARGE/CI0C01492_0006.JPEG.

(150) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J Chem Inf Model* **2015**, *55*, 2562–2574. https://doi.org/10.1021/ACS.JCIM.5B00654.

(151) Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science* **2015**, *2*, 165–193. https://doi.org/10.1007/S40745-015-0040-1.

(152) Miao, Y.; Merz, K. M. Acceleration of High Angular Momentum Electron Repulsion Integrals and Integral Derivatives on Graphics Processing Units. *J Chem Theory Comput* **2015**, *11*, 1449–1462. https://doi.org/10.1021/CT500984T/ASSET/IMAGES/LARGE/CT-2014-00984T_0009.JPEG.

(153) Manathunga, M.; Aktulga, H. M.; Götz, A. W.; Merz, K. M. Quantum Mechanics/Molecular Mechanics Simulations on NVIDIA and AMD Graphics Processing Units. *J Chem Inf Model* **2023**, *63*, 711–717. https://doi.org/10.1021/ACS.JCIM.2C01505/SUPPL_FILE/CI2C01505_SI_002.ZIP.

(154) Manathunga, M.; Miao, Y.; Mu, D.; Götz, A. W.; Merz, K. M. Parallel Implementation of Density Functional Theory Methods in the Quantum Interaction Computational Kernel Program. *J Chem Theory Comput* **2020**, *16*, 4315–4326. https://doi.org/10.1021/ACS.JCTC.0C00290/SUPPL_FILE/CT0C00290_SI_002.ZIP.

(155) Heerdt, G.; Zanotto, L.; Souza, P. C. T.; Araujo, G.; Skaf, M. S. Collision Cross Section Calculations Using HPCCS. In *Ion Mobility-Mass Spectrometry : Methods and Protocols*; Paglia, G., Astarita, G., Eds.; Springer US: New York, NY, 2020; 297–310. https://doi.org/10.1007/978-1-0716-0030-6_19.

(156) Zanotto, L.; Heerdt, G.; Souza, P. C. T.; Araujo, G.; Skaf, M. S. High Performance Collision Cross Section Calculation-HPCCS. *J Comput Chem* **2018**, *39*, 1675–1681. https://doi.org/10.1002/jcc.25199.

(157) Hu, J.; Liu, Z.; Yu, D. J.; Zhang, Y. LS-Align: An Atom-Level, Flexible Ligand Structural Alignment Algorithm for High-Throughput Virtual Screening. *Bioinformatics* **2018**, *34*, 2209. https://doi.org/10.1093/BIOINFORMATICS/BTY081.

(158) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Computational Molecular Science* **2011**, *1*, 826–843. https://doi.org/10.1002/wcms.31.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(159) Anstine, D. and Z. R. and I. O. AIMNet2: A Neural Network Potential to Meet Your Neutral, Charged, Organic, and Elemental-Organic Needs. *ChemRxiv* **2023**.

(160) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci Adv* **2019**, *5*, 6490 - 6498. https://doi.org/10.1126/sciadv.aav6490.

https://doi.org/10.26434/chemrxiv-2023-hc8jv-v3 **ORCID:** https://orcid.org/0000-0001-7981-5162 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0