Moltiverse: Molecular Conformer Generation Using Enhanced Sampling Methods

Mauricio Bedoya[1,2‡], Francisco Adasme-Carreño[1,2*‡], Paula Andrea Peña-Martínez[3], Camila Muñoz-Gutiérrez[1,2], Luciano Peña-Tejo[4], José C. E. Márquez Montesinos[4], Erix W. Hernández-Rodríguez[2,5], Wendy González[4], Leandro Martínez[6], Jans Alzate-Morales[4].*

[1]Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Vicerrectoría de Investigación y Postgrado, Universidad Católica del Maule, Talca, Chile.

[2]Laboratorio de Bioinformática y Química Computacional (LBQC), Departamento de Medicina Traslacional, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile.

[3]Doctorado en Ciencias Agrarias, Facultad de Ciencias Agrarias, Universidad de Talca, Talca, Chile.

[4]Center for Bioinformatics, Simulation and Modeling (CBSM), and Department of Bioinformatics, Faculty of Engineering, Universidad de Talca, Talca 3460000, Chile.

[5]Unidad de Bioinformática Clínica, Centro Oncológico, Facultad de Medicina, Universidad Católica del Maule, Talca 3480094, Chile.

[6]Institute of Chemistry and Center for Computing in Engineering & Science, University of Campinas, Campinas 13083-861 SP, Brazil.

KEYWORDS: Conformer generation, enhanced sampling, ligands, conformational diversity, radius of gyration, cofactors dataset.

ABSTRACT: Accurately predicting the diverse bound-state conformations of small molecules is crucial for successful drug discovery and design, particularly when detailed protein-ligand interactions are unknown. Established tools exist, but efficiently exploring the vast conformational space remains challenging. This work introduces Moltiverse, a novel protocol using enhanced sampling molecular dynamics (MD) simulations for conformer generation. The extended adaptive biasing force (eABF) algorithm combined with metadynamics, guided by a single collective variable (radius of gyration, RDGYR), efficiently samples the conformational landscape of a small molecule. Moltiverse demonstrates comparable accuracy and, in some cases, superior quality when benchmarked against established software like RDKit, CONFORGE, ConfGenX, Torsional diffusion, and Conformator. We present an exhaustive ranking based on eight quantitative metrics and statistical analysis for robust conformer generation algorithms comparison and provide recommendations for their improvement based on our findings. We introduce the *Cofactor*$_{v1}$ dataset, a complementary resource for conformer generator evaluation. Unlike traditional datasets with thousands of single-conformer molecules, the *Cofactor*$_{v1}$ dataset features only seven small molecule cofactors but with hundreds to thousands of experimental conformers *per* molecule (sourced from the PDB). This diversity, encompassing 15-29 rotatable bonds, poses a significant challenge for conformer generation benchmarks. *Cofactor*$_{v1}$ is a complementary dataset that serves as a valuable resource for developing and evaluating conformer generation methods like Moltiverse, pushing the boundaries of accuracy and diversity in this relevant field.

INTRODUCTION

The accurate prediction of the bound-like conformations of small molecules is crucial for many computational methods in drug development, including virtual screening[1–3], structure-based drug design[4–7], molecular docking[8,9], and pharmacophore modeling[10,11]. Exploring the vast conformational space of flexible molecules efficiently is still challenging, especially when the bioactive conformation is unknown or differs significantly from low-energy states. While numerous conformer generation tools exist, there is a pressing need for methods that can balance accuracy, precision, diversity, geometrical correctness, and computational efficiency. The landscape of conformer generation is diverse, encompassing traditional software tools and, more recently, artificial intelligence models designed explicitly for this task. Traditional open-source conformer generators such as Balloon[12], Confab[13], Conforge[14], FROG2[15], OpenBabel[16], and RDKit[17,18] have long been staples in many computational chemistry workflows, offering flexibility and transparency in their methodologies. Commercial tools like CAESAR[19], ConfGen[20], Conformator[21], COSMOS[22], ForceGen[23] and OMEGA[24], often provide highly optimized algorithms[25]. In recent years, this field has seen a surge of innovation with the emergence of artificial intelligence models specifically designed for conformer generation. Approaches such as ConfGF[26], DMCG[27], GeoDiff[28], GeoMol[29], and Torsional diffusion[30,31] have employed advanced machine learning techniques to predict molecular conformations. These AI-driven methods promise to combine the speed of knowledge-based approaches with the accuracy of physics-based methods, potentially revolutionizing the field. Conformer generation algorithms encompass diverse approaches, including systematic and stochastic searches, molecular simulations, distance geometry, knowledge-based methods, AI techniques, and combinations thereof. Systematic search methods such as Conformator[21], which employs a knowledge-based algorithm, systematically

explore the torsional space of molecules, proving effective for compounds with a moderate number of rotatable bonds. However, stochastic sampling techniques often prove more suitable for highly flexible molecules, such as macrocycles[14,21,32], employing random sampling to handle a broader range of molecular flexibilities and ring template libraries[33,34]. The distance geometry (DG) approach implemented in RDKit defines upper and lower distance bounds between all pairs of atoms in a molecule[32,35]. These bounds are typically derived from chemical knowledge, experimental data, and structural constraints. DG approach generates conformations by randomly selecting distances within these bounds for each atom pair, followed by an embedding process that mathematically transforms the interatomic distance matrices into three-dimensional Cartesian coordinates, thereby generating the final molecular conformations. Knowledge-based approaches have emerged as a fast and often accurate alternative to sample every rotatable bond, incorporating libraries of pre-computed torsion angles and fragment conformations derived from experimental data[14,21,36]. For applications demanding high accuracy, force field-based methods[23] offer a physics-based approach using molecular mechanics, while quantum chemical approaches like the Conformer-Rotamer Ensemble Sampling Tool (CREST) method[37] employ semiempirical calculations to guide conformer generation. CREST combines metadynamics simulations with semiempirical calculations employing a history-dependent biasing potential based on atomic RMSD as a collective variable to explore the potential energy surface. Through iterative sampling and geometry optimization, CREST generates conformational ensembles that account for electronic effects. This approach has proven effective for complex molecular systems such as macrocycles[37,38] and metal-containing compounds[39], where conventional force field methods show limitations. However, despite using relatively efficient semiempirical methods, CREST calculations remain computationally demanding and require significantly longer processing times

compared to traditional approaches. In this work, we introduce Moltiverse, a novel approach for conformer generation using enhanced sampling molecular dynamics (MD) simulations to explore the conformational landscape of small molecules. At its core, Moltiverse combines the extended adaptive biasing force algorithm with metadynamics (M-eABF), guided by a single collective variable – the radius of gyration (RDGYR). This unique combination enables efficient and exhaustive generation of conformer ensembles in a short simulation time, overcoming energy barriers that often trap traditional MD simulations in local minimum. Unlike systematic and stochastic approaches, Moltiverse's MD-based method uses a physics-based potential for the exploration of conformational space, which avoids generating potentially fictitious geometries. This approach is particularly advantageous for molecules with complex energy landscapes or those with a high number of rotatable bonds, where traditional methods might struggle to sample relevant conformations efficiently.

MATERIALS AND METHODS.

Dataset Compilation. A search was carried out in the Binding MOAD[40] database, which is enriched for ligand-protein complexes with resolutions less than 2.5 Å extracted from the Protein Data Bank (PDB)[41]. The cofactors with the highest frequency and a number of rotatable bonds greater than or equal to 15 were selected. Then, for each distinct number of rotatable bonds, the cofactor with the largest number of experimental structures was selected. This resulted in 7 different cofactors: Acetyl coenzyme A (**ACO**), Coenzyme A (**COA**), Flavin adenine dinucleotide (**FAD**), Nicotinamide adenine dinucleotide phosphate (**NAP**), Nicotinamide adenine dinucleotide (**NAD**), Guanosine 5'-triphosphate (**GTP**) and Adenosine 5'-triphosphate (**ATP**). The entries in the PDB for these cofactors were downloaded on November 27, 2023, where the number of entries *per* cofactor ranged from 190 (ACO) to 1956 (FAD). These entries were further filtered down by

the following quality conditions: (i) PDB file with validation XML file available, (ii) a resolution of 2.5 Å or less, (iii) an average occupancy value of 1.0, (iv) all heavy atoms resolved, and (v) a real space correlation coefficient (RSCC) greater than or equal to 0.95. Multiple copies of a cofactor from the same PDB entry were included if they matched the above criteria. Covalent molecules were also included.

Post-processing of experimental conformers.

Each selected copy (conformer) of the cofactors was extracted from the original PDB file using the chem.cr library, an open-source programming package implemented in the modern Crystal language[42] for manipulating and analyzing molecular structures from computational chemistry files[43]. The maximum common substructure (MCS) algorithm implemented in the RDKit library[25,44] was used to rearrange the order of the atoms of each conformer to facilitate the Root Mean Square Deviation (RMSD) calculation in subsequent analyses. Both the original and rearranged structures are available for download (refer to the Data and Software Availability section). Structures not recognized by the MCS algorithm were discarded for practical reasons, even if they met the quality criteria defined above.

Moltiverse protocol

Moltiverse is a ligand conformer generator available as an open-source command line application written in the Crystal language[42] based on the chem.cr library[43]. Moltiverse uses the robust ecosystem of open-source applications to process the molecules and perform conformational sampling, using the chem.cr library as its foundation for file conversion, input creation, calculation execution, geometric analysis, and clustering. The conformer generation protocol consists of seven main steps: (i) molecular pre-processing which includes conversion of the SMILES (Simplified

Molecular Input Line Entry System) code into three-dimensional coordinates using Open Babel software[16], (ii) initial conformer stretching, (iii) parameterization of the molecule with the GAFF2[45] force field using AmberTools[46], (iv) energetic minimization, (v) sampling of the molecule conformations with the M-eABF[47] method in vacuum using the NAMD[48] molecular dynamics engine, (vi) structure clustering, and (vii) conformer ensemble refinement using electronic structure optimization calculations with XTB software[49]. Each step of the protocol is described in detail in the following subsections.

Molecular preprocessing. Conformer generators typically use the SMILES code as the input format for molecules of interest. This approach has the advantage of avoiding bias in the initial structure under study. Moltiverse exclusively accepts an SMI file which contains the SMILES code and name of one or more molecules. The conversion into three-dimensional coordinates is done by the Open Babel[16] software. The chem.cr library is used for reading the converted files and their properties, which are useful for the following steps.

Molecular conformation stretching. An initial geometry of the given molecule is generated using Open Babel[16] and the gen3d[50] command in the default "medium" mode, which performs a fast conformer search and geometry optimization with the MMFF94 force field[51]. About 1000 conformations are randomly generated and the most extended (largest radius of gyration) conformation is selected. Starting from an extended conformation helps to achieve greater reproducibility in the conformational sampling step.

Ligand parameterization. The parameters of the molecule for the GAFF2[45] v2.20 force field are automatically generated using the antechamber[52] program from the AmberTools 23[46] suite, where partial charges are obtained with the AM1-BCC charge model.

Energy minimization. Energy minimization of the molecule in vacuum is performed with the NAMD[48,53] molecular simulation engine using the conjugate gradient method[54] up to 5000 cycles. Non-bonding interactions are excluded for bonded atoms 1-3 and electrostatic interactions are modified by the constant factor of 0.83 for bonded atoms 1-4. The van der Waals interactions for bonded atoms 1-4 are divided by 2.0 (scnb=2.0). The cut-off for van der Waals and electrostatic interactions is set to 50 Å to avoid instabilities in large molecules.

Conformational sampling. The eABF method[55] combined with Metadynamics sampling[56] with a single collective variable; the radius of gyration (RDGYR) as implemented in the Colvars module[57] is used to drive the exploration of the conformational landscape of the molecules in vacuum. All ligand atoms are included in the RDGYR collective variable. The calculation is divided into 12 windows (0.5 Å width), spanning from 3 to 9 Å. A M-eABF simulation is run for 2 ns, and 2500 frames equally spaced in time are stored *per* window, yielding a total of 24 ns of simulation and a maximum of 30000 structures for further analysis. Harmonic walls with a force constant of 10 kcal/mol are used to softly restrain the exploration of the molecule within each window. Both the bin width and the standard deviation between the collective variable and the extended degree of freedom (extendedFluctuation parameter), are set to 0.05 Å, and at least 250 samples must be reached in a bin prior to fully applying the bias (FullSamples parameter) in the ABF calculation. The Gaussian hill´s weight is set to 3.0 kcal/mol and the Gaussian hill width is set to 3.0 Å, while a new Gaussian hill is added every 50 steps to the metadynamics potential. Note that these settings have not been tuned for accurate energy estimation, but to force extensive exploration of the conformational space of the studied molecules in a short simulation time. The integration time for each step of the simulation is 1 femtosecond. Template files for minimization

and sampling are available at GitHub: https://github.com/ucm-lbqc/moltiverse (refer to the Data and Software Availability section).

Structure clustering. All molecular structures generated in the conformational sampling are clustered down to the specified value of the number of output conformers based on the all-atom root mean squared deviation (RMSD) using hierarchical clustering with the single linkage rule. Structures are first aligned (best fit) to only account for internal degrees of freedom (no translation and rotation) in the distance metric. The representative of each cluster is defined as the structure with the lowest average RMSD to all the structures in the cluster (centroid). Clustering was performed with the open source hclust library[58] written in the Crystal language.

Conformer ensemble refinement. Due to the applied potential to force extensive sampling, geometric anomalies such as long bonds were observed in the selected structures during development. To avoid this issue, a two-step post-processing refinement was implemented. First, energy minimization using the GAFF2 force field in NAMD is performed (MM), limited to 500 steps to preserve the initial conformations. This is followed by quantum mechanical (QM) optimization using the semi-empirical GFN2-xTB[49,59] method with XTB software, employing the "crude" optimization level with convergence criteria of 5×10^{-4} Eh for energy and 1×10^{-2} Eh·α^{-1} for gradient, and a maximum of 1500 iterations. If QM optimization fails to converge, the MM conformation is retained to avoid excessive computational time. This refinement process effectively corrects geometric issues while maintaining the diversity of the conformer ensemble. The output of this final step is written to an SDF file.

Conformer generation benchmark. The output of Moltiverse was compared to that of five available tools for the seven cofactors collected in the *Cofactor*$_{v1}$ dataset. The (isomeric if

available) SMILES codes of these molecules were obtained from the PubChem[60] database. The functionality for conformer generation implemented in RDKit was used with the ETKDG$_{v3}$ (RDKit$_{etkdg3}$), ETKDG$_{v3}$-mf (RDKit$_{etkdg3\text{-}mf}$) and KDG (RDKit$_{kdg}$) algorithms[36]. CONFORGE was used with the systematic best (CFG$_{sys\text{-}best}$), systematic default (CFG$_{sys\text{-}def}$) and stochastic (CFG$_{stochastic}$)[14] modes. Torsional Diffusion was used with the "drugs" model using default parameters (TD$_{drugs}$)[30] and the particle guidance sampling variant (TD$_{drugs\text{-}pg}$)[31]. Conformator was used with the fast (CFM$_{fast}$) and best (CFM$_{best}$) modes[21]. ConfGenX was used in the default mode with the OPLS3e force field (CFX$_{def\text{-}opls3e}$)[20,61]. A full compilation of abbreviations and their corresponding definitions is available in the Abbreviations section. Moltiverse was used with parameters reported in this cofactor-optimized protocol (Moltiverse$_{c1}$, where 'c1' denotes this first cofactor-specific protocol). Future protocols for other types of molecules (e.g., macrocycles, peptides, etc.) may be implemented under different modes. The twelve algorithm/mode combinations and their related features are summarized in **Table 3**. In all cases, the programs were requested to generate 250 conformers, and ten replicas were run *per* calculation. Detailed configurations and input scripts for ensemble generation are available at GitHub: https://github.com/ucm-lbqc/moltiverse and at DOI: https://doi.org/10.6084/m9.figshare.27346974.v3 (refer to the Data and Software Availability section).

Quality and structural diversity measures. Eight descriptors were calculated to evaluate the conformational diversity of the ensembles and the ability of the algorithms to reproduce the experimental conformations. For every experimental structure collected in the dataset of a given molecule, the best-fit RMSD was computed against all conformers generated by an algorithm and the minimum RMSD value (RMSD$_{min}$) was selected as measure of the accuracy of reproducing

the structure. Therefore, the number of $RMSD_{min}$ values varies from molecule to molecule as they have different numbers of collected structures (**Table 1**). Additionally, the mean of the $RMSD_{min}$ values at and below the 2.5th percentile ($RMSD_{low}$) and the mean of the $RMSD_{min}$ values at and above the 97.5th percentile ($RMSD_{high}$) were calculated as well to obtain an estimate of the lower and upper bounds of the prediction accuracy, respectively. The precision of an algorithm to reproduce experimental structures was defined as the standard deviation of the $RMSD_{min}$ values ($RMSD_{std}$). Conformational diversity was assessed using two complementary approaches: (i) the RDGYR of all the generated conformers by each algorithm, and (ii) the RMSD between the generated conformers. The range of the RDGYR ($RDGYR_{range}$), calculated as the difference between the mean of the RDGYR values at and below the 2.5th percentile and the mean of the RDGYR values at and above the 97.5th percentile values of the generated conformers, determines whether an algorithm can generate compact, extended, or both types of conformations across all the studied molecules. For the $RDGYR_{range}$, a higher value indicates a greater molecular stretching. The average RMSD between all the generated conformers ($RMSD_{matrix}$) indicates whether an algorithm generates highly similar conformations (smaller value) or more diverse structures (larger value). Both $RDGYR_{range}$ and $RMSD_{matrix}$ were also calculated for experimental structures.

The MCS algorithm implemented in RDKit was used to match the common atoms between the generated conformers and the post-processed experimental structures[44], and then the RMSD was calculated using the chem.cr library. The measurements of RDGYR and RMSD considered only the heavy atoms of the molecules.

We also conducted a Spearman's rank (ρ) correlation analysis between the $RMSD_{min}$ and RDGYR values to determine if the different algorithms predict either compact or extended

conformations more accurately. The absolute value of the Spearman correlation |ρ|, denoted as RMSD-RDGYR$_{corr}$, was analyzed independently for each molecule, mode and replica.

Energy evaluation was conducted as an additional assessment criterion. Single point calculations were performed at the B3LYP+D3/6-31G* level of theory[62–66] using the Jaguar software from the Schrödinger suite[67]. Energy values are normalized relative to the lowest energy conformer *per* molecule. This analysis encompassed all conformers generated for the 7 molecules across all algorithms, modes, and replicas, totaling 180,298 single point energy calculations. A small fraction of conformations failed to converge. **Figure S6** and **Table S3** summarize the results, including the distribution of energy and the number of non-converged structures for each method.

Statistical analysis. Seven of the eight variables were evaluated to find statistically significant differences between the ensembles generated by the tested software, and to evaluate reproducibility in the non-deterministic algorithms (Torsional diffusion and Moltiverse). Unless otherwise stated, statistical analysis was carried out on the aggregated values of the different molecules and replicas for each tested algorithm/mode (**Table S1**). For RMSD$_{low}$ and RMSD$_{high}$, all values at and below the 2.5th and at and above the 97.5th percentiles, respectively, instead of the mean value were considered for this purpose. To assess reproducibility, the conformer ensembles of the independent replicas for Torsional Diffusion and Moltiverse were compared by the RMSD$_{min}$ metric for each molecule. Statistical analysis was conducted using the R software version 4.3.2[68]. The data were first assessed for normality[69] and homogeneity of variances[70]. Based on these assessments a Generalized Linear Model (GLM) with Gamma distribution using identity as the link function[71] was employed for every variable. To evaluate the effect of the factors on the data behavior was performed ANOVA[72] in every GLM by type II sum of squares, and according to the results, multiple comparisons were performed using a post hoc analysis with the Tukey

test[73,74]. The significance level (α) for all tests was 0.05, and the significance codes are shown as follows: ***p < 0.001, **p < 0.01, *p < 0.05, ˙p < 0.1, ^{ns}p < 1. **Table S1** provides a summary of the sample sizes for each of the seven variables across all studied modes. **Figure S9** and **Figure S10** display matrices of p-values, categorized by significance codes.

Hardware setup. All the benchmark calculations were performed on a single workstation running a Linux-based operating system with the PopOS 22.04.LTS distribution (kernel version 6.6.6) based on Ubuntu Debian, equipped with an AMD Ryzen 9 - 3950X 16-core processor 3.5 GHz and 16 GB of RAM. Moltiverse and Torsional Diffusion ran in parallel using the 16 cores, while all the other programs used a single thread. The Moltiverse$_{c1}$ implementation requires more computational resources because it runs the computations in parallel, but it does not require a discrete GPU so it can run on any desktop computer or laptop.

RESULTS

Benchmark dataset. The compiled dataset of organic small molecules, referred to as *Cofactor*$_{v1}$ hereinafter, comprises seven biologically significant cofactors chosen for their prevalence and structural complexity. **Table 1** shows the summary of key properties of the dataset for each molecule. The dataset contains hundreds and thousands of high-quality, experimentally determined unique conformers *per* molecule in the bound state. **Figure S1** shows the distribution of the RSCC values and resolutions of the *Cofactor*$_{v1}$ structures, which reveals a relatively homogeneous distribution of these properties across the dataset, indicating consistency in the quality of the experimental structures. The RSCC is a metric that quantifies the agreement between calculated and observed electron density maps[75–77]. The chosen cofactors exhibit a wide range of molecular size and flexibility, where the number of total atoms varies from 47 (ATP) to 89 (ACO),

the number of heavy atoms varies from 31 (ATP) to 53 (FAD), and the number of rotatable bonds range from 15 (ATP) to 29 (ACO). This dataset offers a complementary approach for evaluating conformer generators compared to widely used benchmarks like the Platinum diverse dataset[25], which contains 2859 different ligand structures with 10 or more heavy atoms and 1 to 16 rotatable bonds. The Geometric Ensemble of Molecules (GEOM) dataset[78] comprises two subsets: GEOM-QM9, containing conformers of approximately 133,000 unique molecules, and GEOM-DRUGS, with conformers of about 317,000 unique molecules. The Platinum diverse dataset exhibits an average of 6.6 ± 3.6 rotatable bonds, while GEOM-QM9 has 2.2 ± 1.6, and GEOM-DRUGS shows a similar pattern to the Platinum diverse dataset with 6.5 ± 3.0 rotatable bonds. In contrast, the *Cofactor*$_{v1}$ dataset contains very few but highly flexible molecules (7) and 8412 structures, showing a substantially higher average of 19.9 ± 3.5 rotatable bonds. The *Cofactor*$_{v1}$ dataset intentionally incorporates molecules with many rotatable bonds as it reflects on the understanding that conformer generation becomes increasingly difficult as the number of rotatable bonds increases[14,79], thus providing a more stringent test of the generation capabilities.

Table 1. Structural features of the *Cofactor*$_{v1}$ dataset.

Molecule	Atoms[a]	Heavy atoms	Rotatable bonds[b]	Entries[c]	Structures[d]
ACO	89	51	29	190	132
COA	84	48	28	548	292
FAD	86	53	23	1956	2782
NAP	76	48	20	1374	1776
NAD	71	44	17	1308	2009
GTP	48	32	16	468	634
ATP	47	31	15	897	787
Total/Avg.	74	46	20	6741	8412

[a] Total number of atoms including hydrogens. [b] Rotatable bonds are defined as: single, non-ring bonds between heavy atoms, excluding amide C-N and terminal bonds. [c] Total PDB entries. [d] Total number of experimental conformers *per* molecule.

Ensemble size and accuracy. For all generators, it was requested to generate a maximum of 250 structures for each replica. Due to the characteristics of each algorithm, the ensemble conformations were not of the same size. Table 2. Conformer ensemble sizes generated for each cofactor by the algorithms *per* replica. shows a summary of the number of conformers generated by each algorithm. In the 10 replicates, the algorithms were reproducible in terms of the number of conformers generated. RDKit, Conforge, Torsional diffusion and Moltiverse generated the maximum requested number of conformations in all modes, however, Conformator and ConfGenX generated ensembles with smaller sizes. Although the ensembles are not the same size, the same comparisons were made between the algorithms to measure the accuracy of the calculations and the conformational diversity.

Table 2. Conformer ensemble sizes generated for each cofactor by the algorithms *per* replica.

Algorithm	Mode	Ensemble size						
		ACO	COA	FAD	NAP	NAD	GTP	ATP
RDKit	etkdg3	250	250	250	250	250	250	250
	etkdg3-mf	250	250	250	250	250	250	250
	kdg	250	250	250	250	250	250	250
Conforge	sys-best	250	250	250	250	250	250	250
	sys-def	250	250	250	250	250	250	250
	stochastic	250	250	250	250	250	250	250
ConfGenX	def-opls3e	250	250	128	53	142	192	250
Torsional Diffusion	drugs	250	250	250	250	250	250	250
	drugs-pg	250	250	250	250	250	250	250
Conformator	fast	188	112	96	117	113	43	36
	best	238	215	185	211	211	231	173
Moltiverse	c1	250	250	250	250	250	250	250

In previous studies, the accuracy of conformer algorithms has been defined as the minimum RMSD value between any structure of the generated ensemble, and a single experimental structure for every tested molecule[7,18,25,79]. In general, the mean and median of the minimum RMSD values

(often referred to average minimum RMSD, AMR) and the percentage of structures generated with RMSD values below various thresholds (thresholds) have been reported for a variety of datasets, with the Platinum diverse dataset[25] being the most commonly used. Although this approach provides a global evaluation of an algorithm to handle different types of molecules, it only considers one experimental conformation *per* molecule. Therefore, previous evaluations have not estimated the precision of the prediction and have ignored the ability of the algorithms to generate a diverse ensemble of conformers of a molecule. In our approach, it is possible to evaluate how accurately and precisely the tested algorithms can reproduce experimentally observed ensembles of bound-like conformations. Furthermore, we have assessed the conformational diversity within the generated ensembles of conformers as it has not been reported previously.

Conformer generation tools. Moltiverse was compared against five available tools for conformer generation in the *Cofactor*$_{v1}$ dataset. **Table 3** summarizes the general settings employed by each algorithm and mode. Several algorithms use RMSD-based clustering to select final conformations. Moltiverse also employs clustering to select a subset of conformations from the total generated pool. This approach allows it to return precisely the number of conformations requested by the user, as long as this number doesn't exceed the total conformations generated. Moltiverse allows post-sampling re-clustering to create ensembles of various sizes, a feature currently exclusive to this algorithm. Most algorithms perform geometry optimization using established force fields: UFF[80], MMFF94s[81,82], OPLS3e[61], or GAFF2[45,46,83]. Moltiverse includes an additional QM optimization step using the semi-empirical GFN2-xTB method[59]. Among the algorithms evaluated in this study, most are either open-source and/or available for academic use. ConfGenX is the only algorithm with an exclusively commercial license. Moltiverse uses open-

source and academic libraries and operates under a GPL-3.0 license. This license permits free use for non-commercial purposes and academic research.

Table 3. Conformer generation algorithm settings used for *Cofactor*$_{v1}$ dataset benchmarking.

Algorithm	Mode	Clustering	Force field optimization	QM optimization	Software versión	License
RDKit	Etkdg3	-	UFF[a]	-	RDKit 2022.09.4	CC BY-SA 4.0
	etkdg3-mf	-	MMFF94s[b]	-		
	kdg	-	-	-		
CONFORGE	sys-best	RMSD	MMFF94s_RTOR NO_ESTAT[c]	-	1.0.0	LGPL-2.1
	sys-def	RMSD	MMFF94s_RTOR NO_ESTAT[c]	-		
	stochastic	RMSD	MMFF94s_RTOR[d]	-		
ConfGenX	def-opls3e	-	OPLS3e[e]	-	Schrodinger suite 2023-4	Commercial
Torsional Diffusion	drugs	-	MMFF94s[b]	-	-	MIT
	Drugs-pg	-	MMFF94s[b]	-	-	
Conformator	fast	RMSD	-	-	Unicon 1.4.3	Academic Use and others
	best	RMSD	-	-		
Moltiverse	c1	RMSD	GAFF2[f]	GFN2-xTB (crude)	Moltiverse AmberTools 23 GAFF2 v2.20 NAMD 2.14 XTB 6.6.1 OpenBabel 3.1.0	GPL-3.0 and dependencies licenses

[a]Universal force field (UFF)[80]. [B]MMFF94s parameter set[81]. [c]MMFF94s_RTOR excluding electrostatic interaction terms[82]. [d]MMFF94s using a reparameterization for torsion parameters[82]. [e]OPLS3e force field. The *Cofactor*$_{v1}$ dataset does not have macrocycles. [f]GAFF2 v2.20 available in AmberTools23[46] was used.

Accuracy and precision *per* molecule. The comparison of conformer generation quality metrics between the algorithms is visually depicted in **Figure 1A**. In terms of accuracy, Moltiverse was the best performing algorithm (lowest $RMSD_{min}$ values) for the largest molecule ACO, achieving an $RMSD_{min}$ of 2.20 Å ± 0.2 Å (**Figure S2**, **Table S2**). ConfGenX and $Conformator_{best}$ are slightly worse with $RMSD_{min}$ of 2.34 ± 0.19 Å and 2.42 ± 0.15 Å, respectively. RDKit showed the worst accuracy in any of the three modes tested, where it produced large variability in the $RMSD_{min}$ (low precision). This outcome is more pronounced for larger molecules. The greater number of outliers

in the $RMSD_{min}$ of the ACO molecule evidenced that some experimental structures could not be reproduced by RDKit (**Figure S2, Table S2**). For the second largest molecule, COA, the best generators in terms of $RMSD_{min}$ were Moltiverse, $CONFORGE_{sys\text{-}def}$ and ConfGenX with $RMSD_{min}$ of 2.07 ± 0.19 Å, 2.11 ± 0.26 Å and 2.2 ± 0.19 Å respectively. Same as in ACO, RDKit gave the worse results: 2.81 ± 0.67 Å, 2.75 +- 0.68 Å, and 2.71 +- 0.64 Å for $RDKit_{kdg}$, $RDKit_{etkdg3}$, and $RDKit_{etkdg3\text{-}mf}$ respectively. FAD is the third largest molecule and the one with the largest number of experimental structures (2782). This is a special case where it can be observed that both accuracy and precision are complementary and more valuable than accuracy alone. The three algorithms with the best accuracy without considering precision are: $Conformator_{best}$, $RDKit_{etkdg3}$ and $RDKit_{kdg}$ with $RMSD_{min}$ of 1.74 ± 0.15 Å, 1.92 ± 0.50 Å and 1.97 ± 0.48 Å respectively. Again, RDKit showed the lowest precision, where some structures could not be predicted correctly even within an $RMSD_{min}$ of 4 Å (**Figure S2**). Considering both accuracy and precision, $Conformator_{best}$ still is in the first place, followed by Moltiverse and $Conformator_{fast}$ with $RMSD_{min}$ of 1.99 ± 0.21 Å and 2.04 ± 0.17 Å respectively. Note that for the FAD molecule, the average $RMSD_{min}$ values are beginning to be below 2 Å, which has traditionally been established as the cut-off for acceptable prediction[84–86]. For the NAP molecule, RDKit gives values with higher accuracy, 1.71–1.74 Å, but with a lower precision compared to Conformator and Moltiverse, exhibiting outlier values up to 3.66 Å. The other best algorithms are ordered as: $Conformator_{best}$, $Conformator_{fast}$, $CONFORGE_{sys\text{-}best}$, Moltiverse, $CONFORGE_{sys\text{-}def}$, Torsional diffusion and ConfGenX. For the NAD molecule, the behavior is similar to NAP, where RDKit has a high accuracy but with large outliers, and $CONFORGE_{stochastic}$ has the lowest accuracy ($RMSD_{min}$ of 2.81 ± 0.24 Å). The algorithms with the best combination of accuracy and precision are Moltiverse and $Conformator_{best}$ with $RMSD_{min}$ of 1.81 ± 0.17 Å, and 1.85 ± 0.17 Å respectively. For the

smaller molecules, GTP and ATP, all algorithms reproduced the experimental structures with average $RMSD_{min}$ values below 2 Å, which is expected from how well the algorithms have been calibrated for molecules with few rotatable bonds (<15). In contrast, the sampling based on the RDGYR done by Moltiverse seem to work equally good for both small and large molecules.

Global accuracy and precision. Table 4 summarizes the overall values for the chosen quality metrics, considering all molecules simultaneously. Conformator showed the highest accuracy and precision among the tested algorithms. $RDKit_{etkdg3}$ presented the lowest global $RMSD_{low}$ value (1.21 Å) which indicates the lower bound of the $RMSD_{min}$ values, reflecting the highest accuracy achieved in the dataset. $Conformator_{best}$ and Moltiverse presented the lowest global $RMSD_{high}$ values: 2.14, and 2.16 Å respectively. The $RMSD_{high}$ represents the upper bound of the prediction, and indicates the accuracy in the worst cases. A lower global $RMSD_{high}$ indicates the algorithm's consistency across diverse molecular structures.

In addition, the distributions of the number of experimental conformations predicted at various RMSD thresholds is illustrated in **Figure S8**. For smaller molecules, the distributions are steeper with 100% of the conformers reproduced at an RMSD less than 2 Å. However, these curves shift towards high values as the molecular size increases. Moltiverse demonstrates consistency across all molecules, exhibiting a sigmoid-shaped curve in all cases. In contrast, other algorithms show, at least for one molecule, curves deviating from the sigmoid shape with abrupt steps in their ascent. This pattern suggests that Moltiverse maintains a more uniform behavior across the tested cofactor molecules.

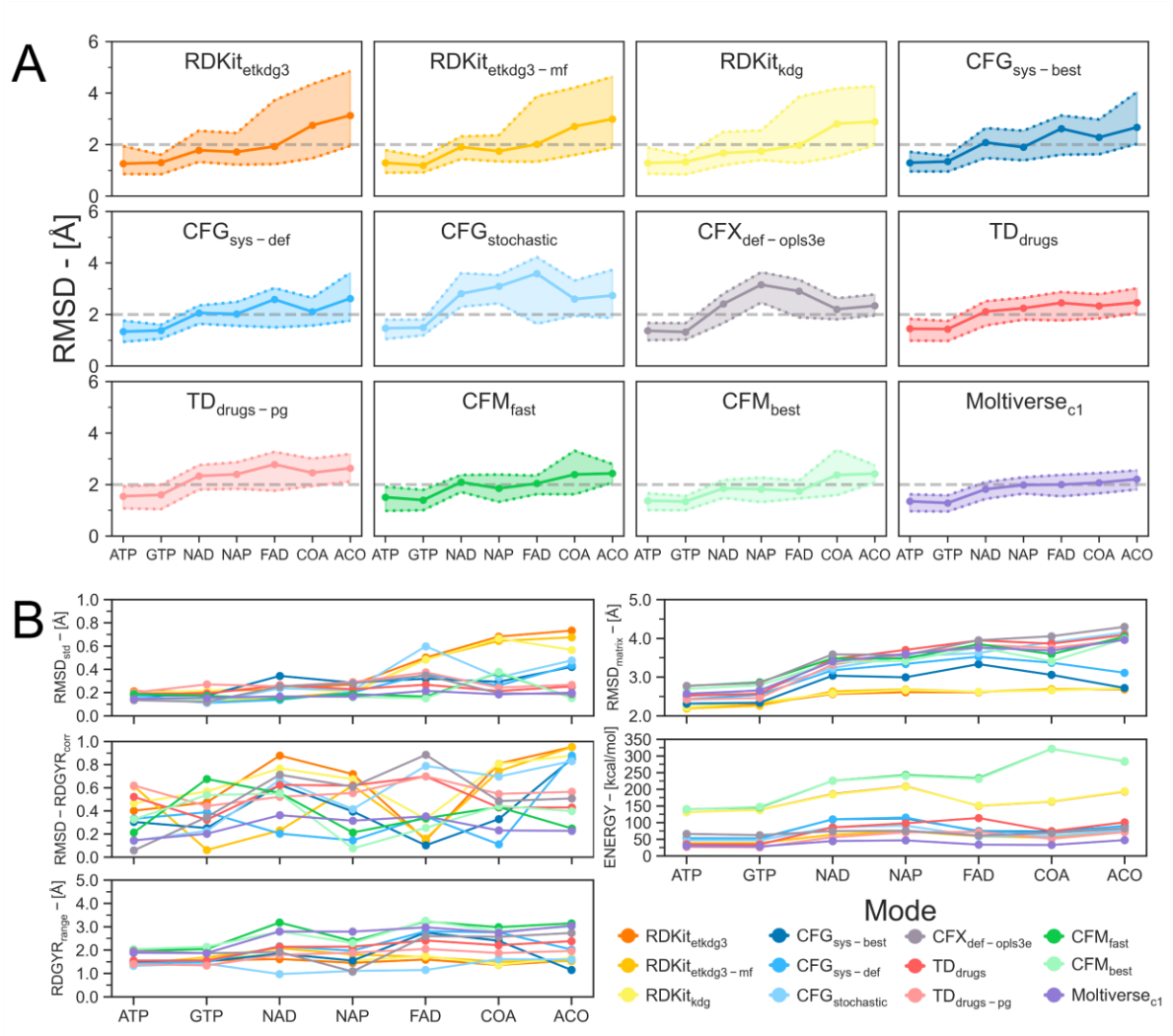

Figure 1. Conformer generation quality comparison for the *Cofactor*$_{v1}$ dataset. Panels are ordered from the molecule with the lowest (ATP) to the highest (ACO) number of rotatable bonds. **A**. The solid line indicates the average $RMSD_{min}$ values, while the lower and upper dotted lines represent the average $RMSD_{low}$ and $RMSD_{high}$ values, respectively. The gray dashed line at 2 Å serves as a reference point. **B**. Only the average of every metric is shown for clarity.

Table 4. Conformer generation quality metrics for the *Cofactor*$_{v1}$ dataset by algorithm.

Algorithm	Mode	$RMSD_{min}$[a]	$RMSD_{std}$[a]	$RMSD_{low}$[a]	$RMSD_{high}$[a]	RMSD - $RDGYR_{corr}$	$RDGYR_{range}$[a]	$RSMD_{matrix}$[a]	ENERGY[b]
RDKit	etkdg3	1.78	0.51	**1.21**	2.89	0.63	1.50	2.53	162.01

	etkdg3-mf	1.84	0.49	1.31	2.84	0.48	1.68	2.54	58.23
	kdg	1.78	0.49	1.23	2.92	0.64	1.60	2.54	162.32
Conforge	sys-best	2.10	0.55	1.43	2.65	0.41	1.81	2.83	78.07
	sys-def	2.12	0.49	1.47	2.52	**0.34***	2.09	3.07	78.07
	stochastic	2.89	0.82	1.89	3.48	0.54	1.31	3.37	66.21
ConfGenX	def-opls3e	2.54	0.67	1.81	2.96	0.51	2.11	3.59	70.28
Torsional Diffusion	drugs	2.15	0.42	1.60	2.56	0.52	2.00	3.45	77.55
	drugs-pg	2.37	0.51	1.68	2.82	0.56	1.71	3.33	53.63
Conformator	fast	1.93	0.31	1.48	2.32	0.38	**2.71***	**3.74**	253.83
	best	**1.75**	**0.28**	1.37	**2.14***	0.37	**2.62***	3.38	229.98
Moltiverse	c1	1.84	0.31	1.45	**2.16***	**0.26***	**2.52***	3.38	**37.79**

[a]Units are Angstroms. [b]Units are kcal/mol. The algorithms with the best overall quality for each accuracy criterion are highlighted in bold. *When multiple algorithms demonstrate statistically equivalent top quality, all such algorithms are marked with asterisks.

Conformational diversity. Conformational diversity was assessed using two approaches, RDGYR range and RMSD matrix measurements for the generated conformers. The RDGYR indicates how compact (smaller value) or extended (larger) is a given structure, and it is particularly useful to evaluate the extent of the RDGYR ($RDGYR_{range}$) to compare the different algorithms (**Figure 1B**). The ensemble generated by RDKit (in any mode) showed a distribution of the RDGYR much narrower than the experimental structures (**¡Error! No se encuentra el origen de la referencia.**), especially for the larger molecules ACO, COA, and FAD. RDKit does not fully cover the experimental RDGYR distribution, which can be related to the poor accuracy of RDKit for these molecules. From these results, it can be suggested that RDKit tends to generate extended molecules. $Conformator_{fast}$, $Conformator_{best}$, and Moltiverse showed the highest global $RDGYR_{range}$ values: 2.71, 2.62, and 2.52 Å, respectively (**Table 4**), which indicates that these algorithms are able to cover a higher diversity of compact and extended molecules. This also shows their potential to generate structures outside the training datasets. However, RDKit, CONFORGE, and Moltiverse produced the most extended structures unlike $Conformator_{fast}$. Although RDKit had the lowest precision in reproducing the structures for FAD ($RMSD_{std}$ of 0.48–0.50 Å), it

presented a higher average accuracy ($RMSD_{min}$). The latter may be related to the ability of RDKit to generate structures in a range of RDGYR closer to the highly-extended experimental structures (**Figure S3**). As RDKit has a tendency for generating extended conformers, they mostly fit the experimental RDGYR range. For the small molecules, GTP and ATP, all generators covered the RDGYR ranges of the dataset, and this is reflected in the better accuracy overall.

The analysis of the RDGYR raises an immediate question: whether some conformer generators tend to correctly produce more compact or extended structures. To answer this question, a correlation analysis was performed between the $RMSD_{min}$ and RDGYR values of the experimental structures. The Spearman's rank correlation (ρ) values are shown in **Figure 2** for all tested algorithms. The correlation plots for the ACO molecule are shown in **Figure 2A** as an example, while the plots for all the molecules are shown in **Figure S5**. RDKit (rightmost panel in **Figure 2A**) shows a strong ($|\rho| \sim 0.9$) negative correlation, meaning that compact molecules (low RDGYR) are often badly predicted (high $RMSD_{min}$). On the contrary, a positive correlation indicates that algorithms such as $CONFORGE_{stochastic}$ ($|\rho| \sim 0.8$) and ConfGenX ($|\rho| \sim 0.5$) are better (low $RMSD_{min}$) at generating compact (low RDGYR) structures for the ACO molecule. Moltiverse and $CONFORGE_{sys\text{-}def}$ achieved the overall smallest |ρ| (**Table 4**), 0.26 and 0.34, respectively, suggesting that both algorithms can generate both compact and extended structures with similar accuracy.

The $RMSD_{matrix}$ has been used as a metric for assessing conformational diversity in studies examining how drug-like molecules change shape upon binding to proteins[87]. This matrix was expressed as the percentage of ensemble structures at various RMSD thresholds. The current study adopts a similar approach by calculating the RMSD within the generated ensembles as measure of structural diversity, but introduces two key modifications: first, the average of the $RMSD_{matrix}$

values (**Figure 1B, Table 4**) is used rather than the percentage distribution, and second, statistical analysis was employed to compare means across different conformer generation algorithms (**Figure S9**). The distribution of the RMSD values in the matrix is shown in **Figure S4**. $Conformator_{fast}$ presented the higher $RMSD_{matrix}$ (3.74 Å), which indicates that it produced the most diverse conformer ensemble. Conversely, the RDKit algorithm, across its various modes, generated conformational ensembles with the least diversity ($RMSD_{matrix}$ between 2.53–2.54 Å). Moltiverse and other algorithms such as $Conforge_{stochastic}$, ConfGenX, and Torsional Diffusion reported ensembles slightly less diverse than $Conformator_{fast}$ ($RMSD_{matrix}$ values between 3.4–3.6 Å).

Energy evaluation. The energy of all conformers in each generated ensemble was evaluated through density functional theory (DFT) calculations to avoid inaccuracies of the classical force fields employed by the different algorithms. This metric aimed to identify the algorithms capable of producing lower energy structures, serving as an indicator of the geometric correctness of the generated molecules. However, this metric should be interpreted cautiously, as experimental structures bound to proteins may exhibit internal energy strain, and although such structures reflect the free energy minimum, those states do not necessarily coincide with the structure of minimum potential energy[84,88]. Nevertheless, excessively high energies could indicate problematic geometries, as previously observed for algorithms such as Balloon, RDKit, Frog2, ConfGen, cxcalc, and MOE[25,79]. Lower energy states suggest sufficiently optimized geometries suitable for subsequent, more precise calculations of molecular properties. The energy distributions are shown in **Figure S6**. The average of the normalized energy (with respect to the lowest-energy state) is indicative of the energy window in a conformer ensemble. Moltiverse demonstrates the lowest global mean normalized energy value (37.79 kcal/mol) among all evaluated algorithms. This

superior result can be attributed to Moltiverse's protocol, which incorporates a conformer refinement step using the semi-empirical GFN2-xTB method that ensures geometric correctness. Torsional Diffusion ranks second, with values of 53.63 and 77.55 kcal/mol for $TD_{drugs\text{-}pg}$ and TD_{drugs} modes, respectively. Torsional Diffusion exhibited some notable high-energy conformations of the larger molecules, specifically NAD, NAP, FAD, COA, and ACO. Further inspection revealed that while the energy calculations converged, some of the geometries were physically unrealistic (**Figure S7**). It is noteworthy that despite employing MMFF94s force field optimization, Torsional Diffusion still generates conformers with a few geometric issues. This observation suggests that the MMFF94s optimization step is insufficient to detect or rectify the unrealistic conformations produced by the AI model. Interestingly, Conformator and RDKit generated conformers with the highest energy among the evaluated algorithms. One might initially assume that these higher energy values correspond to experimental structures due to strain, particularly when considering Conformator's ability to reproduce such structures. However, it is crucial to note that energy calculations encountered two types of failure: those due to unrecognizable topology due to issues in the initial geometries and those resulting from energy convergence issues, also, likely stemming from geometrical errors. **Table S3** quantifies the amount of calculation failures for each molecule. The larger, more complex molecules—NAD, NAP, FAD, ACO, and COA—accounted for most of these failures. Among the evaluated algorithms, RDKit exhibited the highest number of total failures (5870), followed by Conforge (3480), and Conformator (2100). ConfGenX, Torsional Diffusion, and Moltiverse demonstrated significantly fewer failures, with 80, 8, and 4 total failures, respectively. This outcome demonstrates that the quality of the conformers should be considered when evaluating conformer generation, as some

geometric optimization strategies are not enough to ensure structural correctness, which inadvertently diminishes the usefulness of the conformer ensemble.

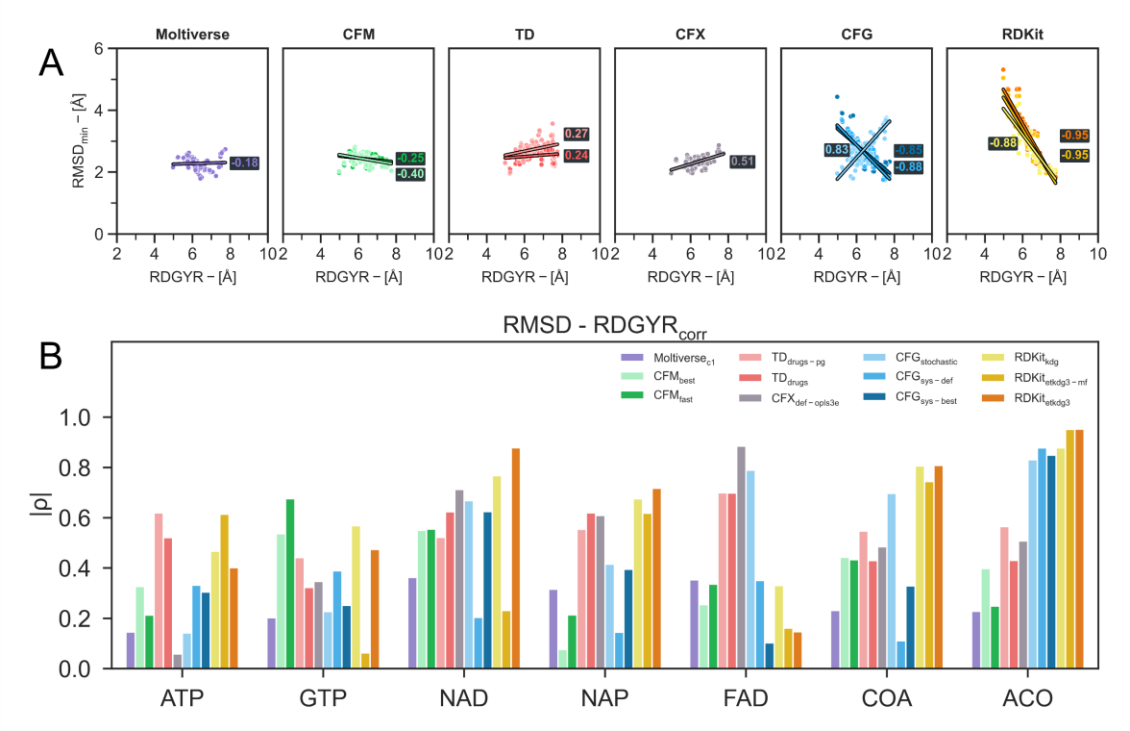

Figure 2. $RMSD_{min}$ vs RDGYR correlation. **A**. Spearman's rank correlation (ρ), data points and trend lines are displayed for the ACO molecule, using data from the first replicate only. The ρ is shown in dark squares. ACO was chosen because it is the largest molecule in the dataset. **B**. Absolute mean of Spearman's rank correlations |ρ|, calculated by computing independent correlations for each replica in each mode.

Systematic quality ranking of conformer generation algorithms. The quality of the conformer generation algorithms was evaluated using eight distinct metrics: $RMSD_{min}$, $RMSD_{std}$, $RMSD_{low}$, $RMSD_{high}$, $RMSD\text{-}RDGYR_{corr}$, $RDGYR_{range}$, $RMSD_{matrix}$, and ENERGY. These metrics collectively assess various aspects of conformational accuracy, diversity, and correctness. Consequently, the following steps were performed to establish a ranking of the tested algorithms based on these metrics:

1. For each metric, the algorithms were ordered from best to worst.
2. Each algorithm was assigned a score based on its position in this order.
3. The scores across all eight metrics were then aggregated to determine an overall quality ranking.
4. The algorithm ranking was refined through statistical analysis. If no statistically significant difference was found between two or more algorithms, they were assigned the same ranking position for each metric.

This aggregate ranking allows for a balanced evaluation that considers multiple facets of the quality of conformer generation. It's important to note that while some algorithms may excel in certain metrics, they might underperform in others. However, it's crucial to consider that the relative importance of each metric may vary depending on the specific application or research context. **Table 5** lists the scores of the algorithms based on their results across each evaluated variable. An ideal algorithm would consistently achieve the top position in all metrics, resulting in a total RANK value of 8. It is important to note that for a more generalized comparison, any additional methods beyond those examined in this study should be evaluated using the same data, metrics, and ranking methodology. According to this ranking, Moltiverse ranks first overall, achieving a RANK of 19. However, it still shows room for improvement in four key areas: accuracy ($RMSD_{min}$), precision ($RMSD_{std}$), lower limit of accuracy ($RMSD_{low}$), and $RMSD_{matrix}$. Conformator, in its best mode, is placed in the second rank with a RANK of 22, demonstrating superior accuracy and precision among the studied algorithms. The third position is shared by $RDKit_{etkdg3\text{-}mf}$ and Torsional Diffusion (TD_{drugs}), both achieving a RANK of 35. $Conforge_{sys\text{-}def}$ and $ConfGenX_{def\text{-}opls3e}$ occupy the fourth and fifth positions, with RANK of 38 and 48, respectively.

Table 5. Conformer generator scores and ranking for the *Cofactor$_{v1}$* dataset by algorithm.

Algorithm	Mode	$RMSD_{min}$	$RMSD_{std}$	$RMSD_{low}$	$RMSD_{high}$	RMSD - $RDGYR_{corr}$	$RDGYR_{range}$	$RSMD_{matrix}$	ENERGY	RANK
Moltiverse	c1	3	3	5	1	1	1	4	1	19
Conformator	best	1	1	4	1	2	1	4	8	22
	fast	4	2	6	2	2	1	1	9	27
RDKit	etkdg3-mf	3	6	3	5	2	3	10	3	35
Torsional Diffusion	drugs	7	4	7	3	3	2	3	6	35
RDKit	kdg	2	5	2	6	3	4	9	7	38
Conforge	sys-def	6	7	6	3	1	2	7	6	38
RDKit	etkdg3	2	8	1	6	3	4	11	7	42
Conforge	sys-best	5	10	5	4	2	3	8	6	43
Torsional Diffusion	drugs-pg	8	9	8	5	3	3	6	2	44
ConfGenX	def-opls3e	9	11	9	7	3	2	2	5	48
Conforge	stochastic	10	12	10	8	3	5	5	4	57

Figure 3 shows the general quality of each algorithm's best mode across all metrics. This visualization is based on a normalization of each variable in a scale from 0 to 1, where 1 represents the most desirable value for each metric. Lower raw values are optimal for $RMSD_{min}$, $RMSD_{std}$, $RMSD_{low}$, $RMSD_{high}$, RMSD-$RDGYR_{corr}$, and ENERGY, while higher raw values are optimal for $RDGYR_{range}$ and $RMSD_{matrix}$. The plot shows how closely the algorithms approach each other in each dimension. A hypothetical ideal algorithm would achieve the optimal position across all metrics, resulting in a polygon that fully extends to the outer limits of the radar plot. It's important to note that the inclusion of additional methods in future comparisons could potentially alter the scale of this graph, as the normalization process relies on the extreme values observed for each metric.

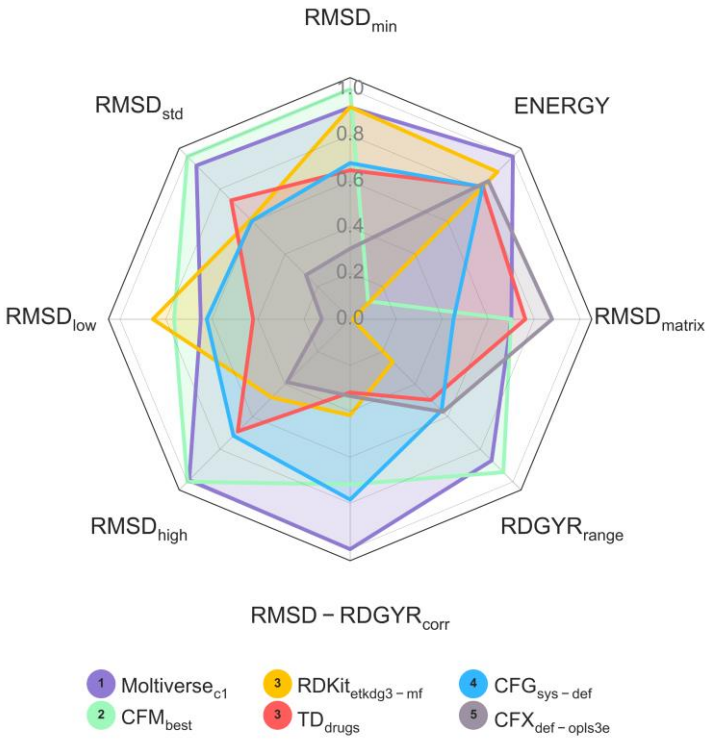

Figure 3. Comparison of the algorithms across eight normalized quality metrics. The radar chart shows normalized values (0-1 scale). For all metrics, values closer to 1 indicate better quality.

To illustrate the best predictions produced by each algorithm, the conformation with the lowest $RMSD_{min}$ along with the corresponding experimental structure is shown in **Figure 4**. Moltiverse achieved the lowest $RMSD_{min}$ values for ACO and COA (1.49 Å and 1.34 Å, respectively). $Conforge_{sys\text{-}def}$ performed best for FAD (0.99 Å), while both Conformator modes produced the best result for NAP (0.77 Å). $TD_{drugs\text{-}pg}$ yielded the lowest $RMSD_{min}$ values for NAD, GTP, and ATP (0.89 Å, 0.51 Å, and 0.58 Å, respectively). These results clearly show that different methods exhibit varying accuracy across molecules, where no single algorithm consistently outperforms others. Notably, Moltiverse was the only method to achieve predictions below 2 Å RMSD for these tested molecules. This result aligns with its superior $RMSD_{high}$ metric, suggesting that even Moltiverse's worst-case predictions are relatively similar to experimental structures.

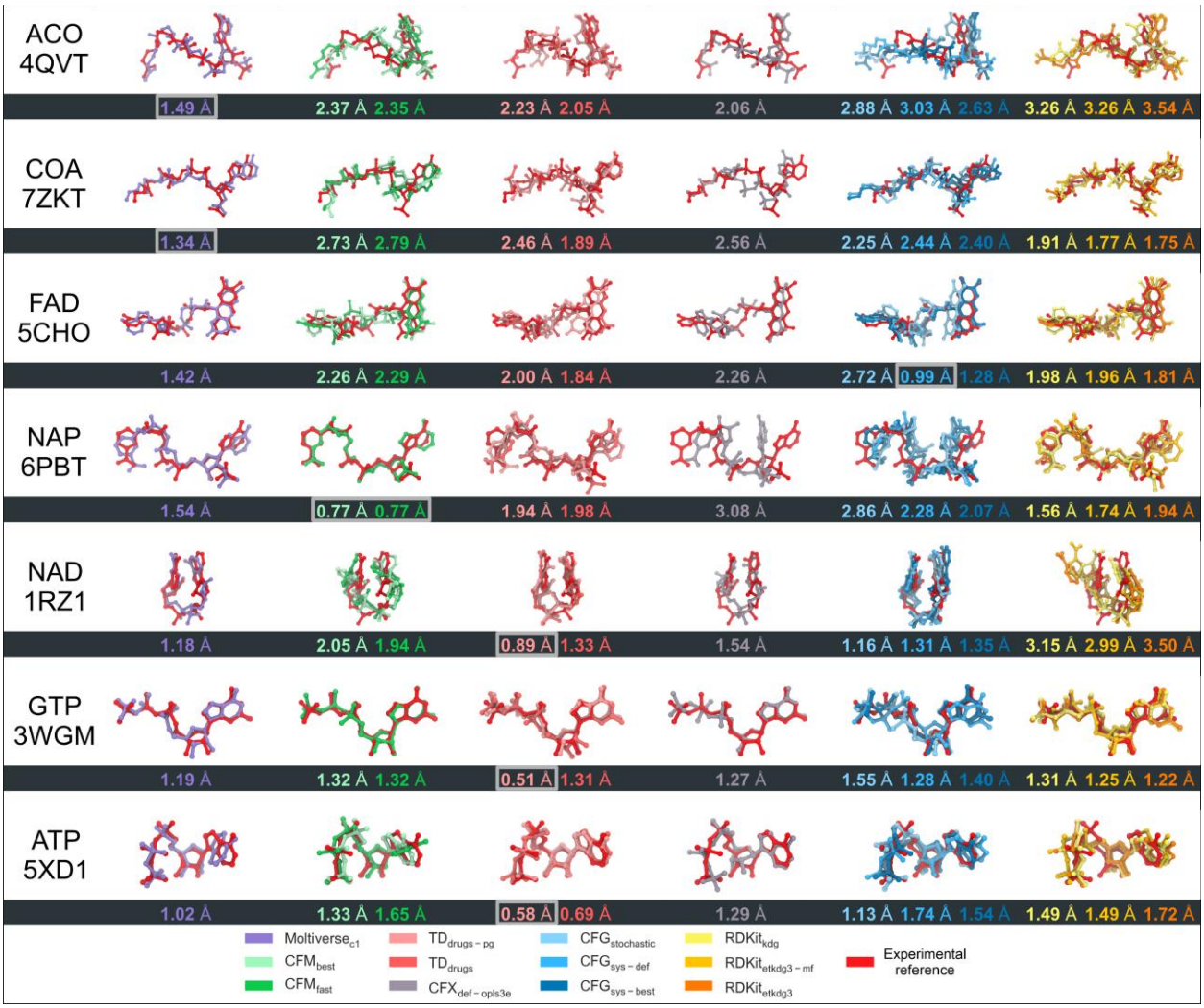

Figure 4. Lowest RMSD conformations for the studied algorithms on the *Cofactor*$_{v1}$ dataset. Each row represents a different molecule labeled with its PDB ID. The columns show conformations generated by the tested algorithms, color-coded as indicated in the legend. The experimental reference structure is shown in red. RMSD values (in Å) are provided below each conformation. The grey boxes enclose the lowest $RMSD_{min}$ value obtained for each molecule among all the algorithms.

Time performance. Moltiverse stands out as the most computationally intensive and time-consuming algorithm, requiring an average of 2254.14 ± 439.98 seconds *per* run on a machine

with the listed specifications (see methods). This is approximately four orders of magnitude slower than the fastest tested algorithm (**Figure 5A**). While Moltiverse and Torsional Diffusion were run using 16 processor cores, the other methods used only a single core, making the performance gap even more pronounced. Conforge emerges as the clear best performing in terms of speed, with its default systematic mode being exceptionally fast. $Conforge_{sys\text{-}def}$ completes in just 0.18 ± 0.06 seconds on average, while $Conforge_{sys\text{-}best}$ takes only slightly longer at 0.23 ± 0.13 seconds. Interestingly, $Conforge_{stochastic}$ is significantly slower at 335.02 ± 164.91 seconds. Conformator offers the next fastest performance after Conforge's systematic modes. Its fast version ($Conformator_{fast}$) averages 28.24 ± 13.19 seconds, while $Conformator_{best}$ takes 36.51 ± 15.66 seconds. This places Conformator as a good middle-ground option between the ultra-fast Conforge systematic modes and slower alternatives. $RDKi_{tkdg}$ takes 37.52 ± 17.96 seconds, $RDKit_{etkdg3}$ 42.95 ± 19.24 seconds, and $RDKit_{etkdg3\text{-}mf}$ is slightly slower with 47.51 ± 20.52 seconds on average. The ConfGenX and the Torsional Diffusion are slower, with times ranging from about 90 to 375 seconds. **Figure 5B** shows the average runtime for each step of the Moltiverse protocol. The MD-based sampling stage is the most time-consuming stage, requiring 1321.05 ± 297.82 seconds, followed by the ligand parameterization at 567.81 ± 384.23 seconds, which is particularly high for larger molecules. These two steps collectively account for approximately 84% of the total processing time. Other stages include clustering (192.06 ± 65.55 s), MM refinement (78.98 ± 15.41 s), and QM refinement (72.99 ± 21.81 s). In contrast, stages such as minimization, molecule processing, and molecular stretching consume significantly less time, each taking less than 21 seconds on average.

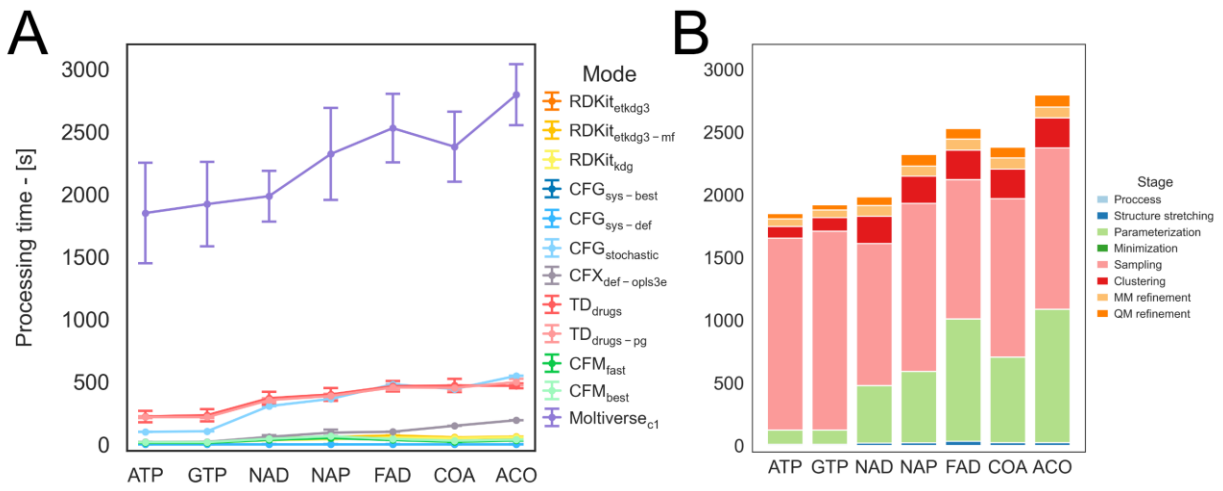

Figure 5. Time performance. **A.** For all generators except Moltiverse and Torsional Diffusion, a single thread was used to perform the calculations. Moltiverse and Torsional Diffusion used the full 16-core processor. **B.** Time *per* stage in the Moltiverse protocol.

DISCUSSION.

The success of Moltiverse with larger, more flexible molecules suggests that other algorithms could improve by incorporating the RDGYR as a variable for conformational sampling or systematic filtering, and thus ensure an improved space exploration when dealing with highly flexible molecules. Most of the tested algorithms exhibit high RMSD-RDGYR$_{corr}$ values, indicating a bias towards either extended structures (as seen with RDKit) or compact structures (Conforge$_{stochastic}$, ConfGenX, and Torsional Diffusion). Indeed, the RDGYR would be beneficial as a metric to quantify the coverage of the conformational space in terms of molecular stretching. This approach would likely lead to more diverse and representative conformer ensembles. RDKit conformers exhibited the highest failure rate in energy calculations hinting at structural issues (**Table S3**). Notably, the RDKit$_{etkdg3-mf}$ mode showed significantly fewer failures compared to the other two RDKit modes. The incorporation of MMFF94s force field minimization in RDKit$_{etkdg3-}$

$_{mf}$ results in lower mean energy values than other RDKit modes, at a relatively modest computational cost. This observation suggests that the integration of optimization based on force fields or QM methods could further enhance the consistency and quality of generated conformers across various conformer generation algorithms.

CONCLUSIONS

Moltiverse, a novel protocol based on open-source applications and the M-eABF method for free energy calculations, demonstrates exceptional capability in generating high-quality, bound-like structures across diverse molecules. The *Cofactor*$_{v1}$ dataset has been introduced to assess the capability of Moltiverse for large molecules. This dataset encompasses multiple high-quality experimental structures for seven highly flexible molecules, complementing existing conformer datasets. Thanks to the significantly larger degrees of freedom to be explored due to an average of 19.9 ± 3.5 rotatable bonds *per* molecule, Cofactor$_{v1}$ represents a challenging benchmark for evaluating conformer generation algorithms. Moltiverse uses the radius of gyration as the single collective variable to drive conformational sampling, which is shown to be particularly effective for large, flexible molecules. Comparative analysis reveals Conformator's superior accuracy and precision among studied algorithms. Moltiverse generates conformers that are often relatively close to known experimental measures (low RMSD$_{high}$) and it reports highly diverse ensembles (high RDGYR$_{range}$), matching Conformator's results, while producing both compact and extended conformations with the same quality (low RMSD-RDGYR$_{corr}$) alongside Conforge. Notably, Moltiverse mostly generates structurally correct low-energy conformers, unlike other algorithms that introduce geometric distortions in the reported structures. However, Moltiverse currently requires greater computational resources compared to other tested algorithms, indicating a need

for optimization in future versions. The proposed holistic evaluation offers a robust framework for assessing conformer generation that could be extended to existing datasets. Moltiverse's capabilities highlight its potential as a valuable tool in computational chemistry and drug discovery, particularly when working with complex, flexible molecules.

DATA AND SOFTWARE AVAILABILITY

All materials for reproducing and extending this study are publicly available through the following public repository: DOI: https://doi.org/10.6084/m9.figshare.27346974.v3. The dataset includes the *Cofactor*$_{v1}$ dataset of original and processed molecules in PDB format, the raw results (CSV files), input file for conformer generation with all algorithms, SMILES codes, Jupyter Notebooks to carried out the benchmark analysis, and R scripts to perform statistical analysis. the Moltiverse source code is available at GitHub: https://github.com/ucm-lbqc/moltiverse

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge at https: future journal link

Tables S1 – S3, Figures S1 – S10.

AUTHOR INFORMATION

Corresponding Author

Mauricio Bedoya – Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Vicerrectoría de Investigación y Postgrado, Universidad Católica del Maule, Talca, Chile. Laboratorio de Bioinformática y Química Computacional (LBQC), Departamento de Medicina

Traslacional, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile; https://orcid.org/0000-0002-3542-7528;

Email:

Francisco Adasme-Carreño – Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Vicerrectoría de Investigación y Postgrado, Universidad Católica del Maule, Talca, Chile. Laboratorio de Bioinformática y Química Computacional (LBQC), Departamento de Medicina Traslacional, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile; https://orcid.org/0000-0003-0971-9081;

Email:

Authors

Paula Andrea Peña-Martínez – Doctorado en Ciencias Agrarias, Facultad de Ciencias Agrarias, Universidad de Talca, Talca, Chile; https://orcid.org/0000-0002-6609-2572.

Camila Muñoz-Gutiérrez – Centro de Investigación de Estudios Avanzados del Maule (CIEAM), Vicerrectoría de Investigación y Postgrado, Universidad Católica del Maule, Talca, Chile. Laboratorio de Bioinformática y Química Computacional (LBQC), Departamento de Medicina Traslacional, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile; https://orcid.org/0000-0003-0919-6903.

Luciano Peña-Tejo – Center for Bioinformatics, Simulation and Modeling (CBSM), and Department of Bioinformatics, Faculty of Engineering, Universidad de Talca, Talca 3460000, Chile; https://orcid.org/0009-0002-2604-7694.

José C. E. Márquez-Montesinos – Center for Bioinformatics, Simulation and Modeling (CBSM), and Department of Bioinformatics, Faculty of Engineering, Universidad de Talca, Talca 3460000, Chile; https://orcid.org/0000-0002-4682-1969.

Erix W. Hernández-Rodríguez – Laboratorio de Bioinformática y Química Computacional (LBQC), Departamento de Medicina Traslacional, Facultad de Medicina, Universidad Católica del Maule, Talca, Chile. Unidad de Bioinformática Clínica, Centro Oncológico, Facultad de Medicina, Universidad Católica del Maule, Talca 3480094, Chile; https://orcid.org/0000-0002-9231-7552.

Wendy González – Center for Bioinformatics, Simulation and Modeling (CBSM), and Department of Bioinformatics, Faculty of Engineering, Universidad de Talca, Talca 3460000, Chile; https://orcid.org/0000-0002-7535-6883.

Leandro Martínez – Institute of Chemistry and Center for Computing in Engineering & Science, Universidade Estadual de Campinas (UNICAMP) 13083-861 SP, Brazil; https://orcid.org/0000-0002-6857-1884.

Jans Alzate-Morales – Center for Bioinformatics, Simulation and Modeling (CBSM), and Department of Bioinformatics, Faculty of Engineering, Universidad de Talca, Talca 3460000, Chile; https://orcid.org/0000-0001-9624-7849.

Author Contributions

M.B and F.A-C implemented the Moltiverse code. P.A.P-M performed the statistical analysis. C.M-G, L.P-T, and J.C.E.M-M performed benchmark analysis under supervision of M.B, F.A-C, and W.G. M.B, F.A-C, E.W.H-R, J.A-M, and L.M conceptualized the work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Funding Sources

M.B. acknowledges FONDECYT - ANID for his postdoctoral grant Nº 3210774. F. A.-C. acknowledges funding from ANID Convocatoria Nacional Subvención a Instalación en la Academia, Convocatoria año 2021, Folio SA772100091. P.P-M. acknowledges the ANID

National Doctorate Scholarship 2019 Folio N° 21190245. L.P-T. acknowledges the ANID National Doctorate Scholarship Folio N° 21231159. E.W.H-R acknowledges funding from his ANID FONDECYT projects N° 3170107 and 11230033. J.A-M. thanks for the financial support from ANID through project FONDECYT N° 1230999. L. M. acknowledges the support of Fapesp (Grants 2018/24293-0 and CNPq (Grant 302332/2016-2).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors thank the software access and computer time granted on the high-performance computing cluster at the Laboratorio de Bioinformática y Química Computacional (LBQC) at Universidad Católica del Maule, Chile. The authors thank the CONFORGE[14] developers for making available the input files for the CONFORGE, Conformator, and RDKit algorithms that were modified and used in this work.

ABBREVIATIONS

$CONFORGE_{sys\text{-}best}$, or $CFG_{sys\text{-}best}$, Conforge systematic best mode; $CONFORGE_{sys\text{-}def}$, or $CFG_{sys\text{-}def}$, Conforge systematic default mode; $CONFORGE_{stochastic}$, or $CFG_{stochastic}$, Conforge stochastic mode; $RDKit_{etkdg3}$, RDKit with the $ETKDG_{v3}$ mode; $RDKit_{etkdg3\text{-}mf}$, RDKit with the $ETKDG_{v3}$-mf mode; $RDKit_{kdg}$, RDKit with the KDG mode; TD_{drugs}, Torsional Diffusion with the drugs default model with the default parameters; $TD_{drugs\text{-}pg}$, Torsional Diffusion with the particle guidance sampling; $Conformator_{fast}$, or CFM_{fast}, Conformator with the fast mode; $Conformator_{best}$, or CFM_{best}, Conformator with the best mode. $CFX_{def\text{-}opls3e}$, ConfGenX in the default mode with the OPLS3e force field.

REFERENCES

1. Patrick Walters, W., Stahl, M. T. & Murcko, M. A. Virtual screening - An overview. *Drug Discov. Today* **3**, 160–178 (1998).

2. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).

3. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **3**, 935–949 (2004).

4. Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discov. Today* **7**, 1047–1055 (2002).

5. Anderson, A. C. The Process of Structure-Based Drug Design. *Chem. Biol.* **10**, 787–797 (2003).

6. Yu, W. & MacKerell, A. D. Computer-Aided Drug Design Methods. in *Methods in Molecular Biology* vol. 1520 85–106 (Humana Press, New York, NY, 2017).

7. McNutt, A. T. *et al.* Conformer Generation for Structure-Based Drug Design: How Many and How Good? *J. Chem. Inf. Model.* **63**, 6598–6607 (2023).

8. Chen, Y. C. Beware of docking! *Trends Pharmacol. Sci.* **36**, 78–95 (2015).

9. Morris, G. M. & Lim-Wilby, M. Molecular Docking. in *Methods in Molecular Biology* vol. 443 365–382 (Humana Press, 2008).

10. Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and

recent advances. *Drug Discov. Today* **15**, 444–450 (2010).

11. Schwab, C. H. Conformations and 3D pharmacophore searching. *Drug Discov. Today Technol.* **7**, e245–e253 (2010).

12. Vainio, M. J. & Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *J. Chem. Inf. Model.* **47**, 2462–2474 (2007).

13. O'Boyle, N. M., Vandermeersch, T., Flynn, C. J., Maguire, A. R. & Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *J. Cheminform.* **3**, 8 (2011).

14. Seidel, T., Permann, C., Wieder, O., Kohlbacher, S. M. & Langer, T. High-Quality Conformer Generation with CONFORGE: Algorithm and Performance Assessment. *J. Chem. Inf. Model.* **63**, 5549–5570 (2023).

15. Miteva, M. A., Guyon, F. & Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **38**, W622–W627 (2010).

16. O'Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).

17. Landrum, G. RDKit: Open-source cheminformatics. https://www.rdkit.org. https://zenodo.org/records/3732262 (2010) doi:10.5281/ZENODO.3732262.

18. Ebejer, J. P., Morris, G. M. & Deane, C. M. Freely available conformer generation methods: How good are they? *J. Chem. Inf. Model.* **52**, 1146–1158 (2012).

19. 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.* **47**, 1923–1932 (2007).

20. Watts, K. S. *et al.* ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **50**, 534–546 (2010).

21. Friedrich, N. O. *et al.* Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **59**, 731–742 (2019).

22. Sadowski, P. & Baldi, P. Small-molecule 3D structure prediction using open crystallography data. *J. Chem. Inf. Model.* **53**, 3127–3130 (2013).

23. 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.* **31**, 419–439 (2017).

24. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).

25. Friedrich, N. O. *et al.* Benchmarking Commercial Conformer Ensemble Generators. *J. Chem. Inf. Model.* **57**, 2719–2728 (2017).

26. Shi, C., Luo, S., Xu, M. & Tang, J. Learning Gradient Fields for Molecular Conformation Generation. in *Proceedings of Machine Learning Research* vol. 139 9558–9568 (ML Research Press, 2021).

27. Zhu, J. *et al.* Direct Molecular Conformation Generation. *Trans. Mach. Learn. Res.* (2022).

28. Xu, M. *et al.* GEODIFF: A GEOMETRIC DIFFUSION MODEL FOR MOLECULAR

CONFORMATION GENERATION. in *ICLR 2022 - 10th International Conference on Learning Representations* (2022).

29. Ganea, O. E. *et al.* GEOMOL: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. in *Advances in Neural Information Processing Systems* vol. 17 13757–13769 (2021).

30. Jing, B., Corso, G., Chang, J., Barzilay, R. & Jaakkola, T. Torsional Diffusion for Molecular Conformer Generation. in *Advances in Neural Information Processing Systems* vol. 35 (2022).

31. Corso, G., Xu, Y., de Bortoli, V., Barzilay, R. & Jaakkola, T. PARTICLE GUIDANCE: NON-I.I.D. DIVERSE SAMPLING WITH DIFFUSION MODELS. in *12th International Conference on Learning Representations, ICLR 2024* (International Conference on Learning Representations, ICLR, 2024).

32. 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.* **60**, 2044–2058 (2020).

33. Kolodzik, A., Urbaczek, S. & Rarey, M. Unique ring families: A chemically meaningful description of molecular ring topologies. *J. Chem. Inf. Model.* **52**, 2013–2021 (2012).

34. Flachsenberg, F., Andresen, N. & Rarey, M. RingDecomposerLib: An Open-Source Implementation of Unique Ring Families and Other Cycle Bases. *J. Chem. Inf. Model.* **57**, 122–126 (2017).

35. Blaney, J. M. & Dixont, J. S. Distance Geometry in Molecular Modeling. in *Reviews in Computational Chemistry* vol. 5 299–335 (John Wiley & Sons, Ltd, 2007).

36. Riniker, S. & Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **55**, 2562–2574 (2015).

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

38. Grambow, C. A., Weir, H., Cunningham, C. N., Biancalani, T. & Chuang, K. V. CREMP: Conformer-rotamer ensembles of macrocyclic peptides for machine learning. *Sci. Data* **11**, 1–9 (2024).

39. Bursch, M., Hansen, A., Pracht, P., Kohn, J. T. & Grimme, S. Theoretical study on conformational energies of transition metal complexes. *Phys. Chem. Chem. Phys.* **23**, 287–299 (2021).

40. Smith, R. D. *et al.* Updates to Binding MOAD (Mother of All Databases): Polypharmacology Tools and Their Utility in Drug Repurposing. *J. Mol. Biol.* **431**, 2423–2433 (2019).

41. Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

42. Crystal: A language for humans and computers. Manas Tech, Buenos Aires, Argentina. https://crystal-lang.org/.

43. Adasme-Carreño, F. chem.cr: Library for Dealing with Computational Chemistry Files. at

https://github.com/franciscoadasme/chem.cr (2024).

44. Landrum, G. *et al.* rdkit/rdkit: 2022_09_4 (Q3 2022) Release. at https://doi.org/10.5281/zenodo.7541264 (2023).

45. Vassetti, D., Pagliai, M. & Procacci, P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J. Chem. Theory Comput.* **15**, 1983–1995 (2019).

46. Case, D. A. *et al.* AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).

47. Fu, H. *et al.* Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *J. Phys. Chem. Lett.* **9**, 4738–4745 (2018).

48. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

49. Bannwarth, C. *et al.* Extended tight-binding quantum chemistry methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* vol. 11 e1493 at https://doi.org/10.1002/wcms.1493 (2021).

50. Yoshikawa, N. & Hutchison, G. R. Fast, efficient fragment-based coordinate generation for Open Babel. *J. Cheminform.* **11**, 49 (2019).

51. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **17**, 490–519 (1996).

52. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type

perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).

53. Kalé, L. *et al.* NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **151**, 283–312 (1999).

54. Hestenes, M. R. & Stiefel, E. Methods of conjugate gradients for solving linear systems. *undefined* **49**, 409 (1952).

55. Fu, H., Shao, X., Chipot, C. & Cai, W. Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. *J. Chem. Theory Comput.* **12**, 3506–3513 (2016).

56. Fu, H., Shao, X., Cai, W. & Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **52**, 3254–3264 (2019).

57. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).

58. Adasme-Carreño, F. hclust: Fast hierarchical clustering algorithms in pure Crystal. at https://github.com/franciscoadasme/hclust (2024).

59. 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.* **15**, 1652–1671 (2019).

60. Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373–D1380 (2023).

61. Roos, K. *et al.* OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **15**, 1863–1874 (2019).

62. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

63. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).

64. Vosko, S. H., Wilk, L. & Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **58**, 1200–1211 (1980).

65. 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.* **98**, 11623–11627 (1994).

66. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, (2010).

67. Bochevarov, A. D. *et al.* Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **113**, 2110–2142 (2013).

68. R Core Team. R: A Language and Environment for Statistical Computing. at https://www.r-project.org/ (2018).

69. SHAPIRO, S. S. & WILK, M. B. An analysis of variance test for normality (complete

samples). *Biometrika* **52**, 591–611 (1965).

70. Brown, M. B. & Forsythe, A. B. Robust Tests for the Equality of Variances. *J. Am. Stat. Assoc.* **69**, 364 (1974).

71. Smyth, G. K. Generalized Linear Models with Varying Dispersion. *J. R. Stat. Soc. Ser. B* **51**, 47–60 (1989).

72. Langsrud, Ø. ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Stat. Comput.* **13**, 163–167 (2003).

73. Tukey, J. W. Comparing Individual Means in the Analysis of Variance. *Biometrics* **5**, 99 (1949).

74. Piepho, H. P. An algorithm for a letter-based representation of all-pairwise comparisons. *J. Comput. Graph. Stat.* **13**, 456–466 (2004).

75. Tickle, I. J. Statistical quality indicators for electron-density maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 454–467 (2012).

76. Smart, O. S. *et al.* Validation of ligands in macromolecular structures determined by X-ray crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 228–236 (2018).

77. Shao, C. *et al.* Simplified quality assessment for small-molecule ligands in the Protein Data Bank. *Structure* **30**, 252-262.e4 (2022).

78. Axelrod, S. & Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **9**, 185 (2022).

79. Friedrich, N. O. *et al.* High-Quality Dataset of Protein-Bound Ligand Conformations and Its Application to Benchmarking Conformer Ensemble Generators. *J. Chem. Inf. Model.* **57**, 529–539 (2017).

80. Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).

81. Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **20**, 720–729 (1999).

82. Wahl, J., Freyss, J., von Korff, M. & Sander, T. Accuracy evaluation and addition of improved dihedral parameters for the MMFF94s. *J. Cheminform.* **11**, 53 (2019).

83. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

84. Peach, M. L., Cachau, R. E. & Nicklaus, M. C. Conformational energy range of ligands in protein crystal structures: The difficult quest for accurate understanding. *J. Mol. Recognit.* **30**, e2618 (2017).

85. Yusuf, D., Davis, A. M., Kleywegt, G. J. & Schmitt, S. An Alternative Method for the Evaluation of Docking Performance: RSR vs RMSD. *J. Chem. Inf. Model.* **48**, 1411–1422 (2008).

86. Vieth, M., Hirst, J. D., Kolinski, A. & Brooks, C. L. Assessing energy functions for flexible docking. *J. Comput. Chem.* **19**, 1612–1622 (1998).

87. 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.* **47**, 2499–2510 (2004).

88. Thompson, H. P. G. & Day, G. M. Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chem. Sci.* **5**, 3173–3182 (2014).