JCTC Journal of Chemical Theory and Computation

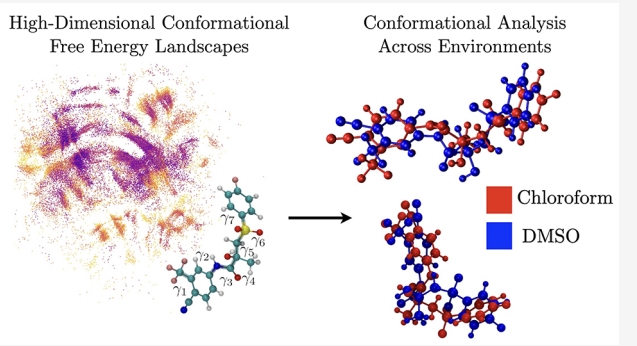
pubs.acs.org/JCTC Article

Sampling High-Dimensional Conformational Free Energy Landscapes of Active Pharmaceutical Ingredients

Alexandre Ferreira, Rui Guo, Ivan Marziano, and Matteo Salvalaglio*

Cite This: J. Chem. Theory Comput. 2025, 21, 12466−12480

ABSTRACT: We present a gridless framework for computing high-dimensional conformational free energy surfaces (FES) of flexible molecules using enhanced sampling trajectories. By combining concurrent well-tempered metadynamics with Density Peaks Advanced (DPA) clustering, our approach bypasses the dimensionality limitations of conventional grid-based FES reconstruction. Free energies are assigned on a per-configuration basis via local density estimation and Zwanzig reweighting, allowing for a direct, resolution-independent mapping of the conformational ensemble. Conformers are identified as density peaks in torsional angle space, and convergence is assessed via systematic consistency metrics. We validate this approach by reproducing the paradigmatic FES of alanine dipeptide and extend it to explore molecules with 4-, 7-, and 11-dimensional torsional angle spaces. As a key application, we investigate the solvent-dependent conformational preferences of bicalutamide in vacuum, chloroform, and DMSO. The predicted global minima reflect the known solvent-induced conformational shift between open and closed forms, in agreement with NMR and crystallographic data. These results demonstrate that our workflow provides a scalable route to high-dimensional conformational free energy landscapes, with direct relevance for polymorphism, solvation, and drug design.



INTRODUCTION

The vast majority of active pharmaceutical ingredients (APIs) are highly flexible molecules, capable of changing shape readily in biological environments, as well as displaying conformational polymorphism in the solid state.[1,2] This polymorphism is particularly important to the drug formulation and manufacturing process, as the pharmacological properties of APIs can depend very strongly on the crystal's polymorphism.[3] As such, quantitatively characterizing the conformational landscape of APIs is of great interest. For instance, it is known that the environment in which a molecule is found impacts its conformational landscape[4−6] however, a systematic approach for mapping, understanding, and quantifying the effects of the environment on the conformational landscape of APIs is still lacking.[7] This gap is partly due to the inherent high-dimensionality of these spaces,[8] which renders their exploration and rationalization challenging, especially when it becomes necessary to explicitly account for the impact of the environment in which an API molecule is found. Here, we present a workflow that combines enhanced sampling molecular dynamics techniques with density-based clustering to simultaneously explore the conformational free energy landscape of flexible API, explicitly accounting for the effects of the environment, and obtain an estimate of the relative free energy of the conformers discovered. These two aspects differentiate our proposed approach from methods based on locally minimizing the potential energy of isolated molecules, where entropic contributions are typically included *a-posteriori* and usually do not include the configurational contribution associated with an explicitly represented environment.[9−12]

Before delving into the details of the method, we introduce the definition of conformational space adopted in this work. A molecule's conformation can be defined using the values of the molecule's freely rotatable dihedral angles[1,6] (henceforth referred to as torsions). The conformational space is, therefore, always bounded and periodic in all dimensions. Veber's rules,[13] a set of heuristics initially designed to predict whether a molecular structure would possess pharmacological properties, can be used to identify the relevant torsions in any given molecular structure. Characterizing conformers through the values of a set of torsions is not without precedent, and several approaches are based on this definition. For example, Torsiflex[14] is a software package that aims to explore a single molecule's potential energy surface

Received: July 28, 2025
Revised: November 11, 2025
Accepted: November 12, 2025
Published: December 4, 2025

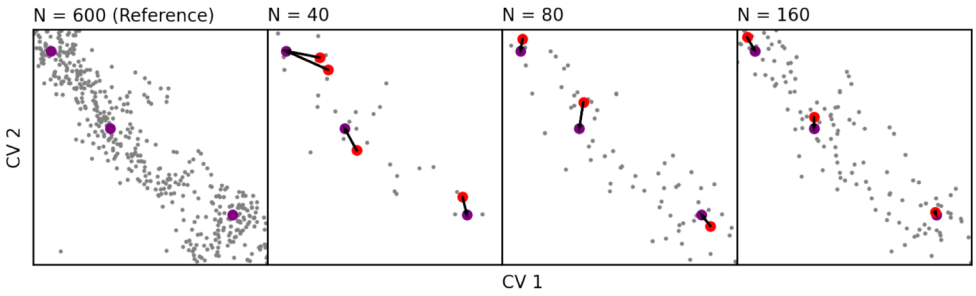


Figure 1. Sketch illustrating the pairing procedure used by the consistency metrics to compare cluster-sets generated from data sets of different sizes. The configurations shown are drawn from the simulation of alanine dipeptide, but the principle illustrated is general. The cluster centers obtained from the largest amount of data form the reference set, shown in purple. Cluster centers generated from smaller amounts of data (shown in red) are paired to the nearest cluster center in the reference set, allowing comparison of distances and energy differences between cluster centers. It is expected that as the data set size grows, the positions and energies of the cluster centers will converge, as seen in Figure 2g,h.

utilizing a semirandom exploration of conformational spaces defined by torsions.

Molecular dynamics (MD) simulations can be used to sample the conformational space and map out the conformational FES of a given molecule. Figure 1 shows different extents of MD sampling of the same collective variable (CV) space. Sampling the probability distribution with MD offers several key advantages; MD sampling is inherently physics-based and enables simulating a molecule in various environments and conditions. The physics-based nature of the sampling means the distribution sampled in conformational space by the MD simulation can be converted into a FES[15] through the relationship:

$$F(\mathbf{S}) = -k_B T \ln p(\mathbf{S}) \quad (1)$$

where $\mathbf{S}$ represents the dihedral angle space $\mathbf{S} = [\gamma_1, \gamma_2, \cdots, \gamma_D]$, where γ_i is one of the D torsional angles of a given molecule. $p(\mathbf{S})$ is the probability distribution in the conformational space, $F(\mathbf{S})$ is therefore the FES in conformational space $\mathbf{S}$, k_B is Boltzmann's constant, and T is the temperature.

To compute $F(\mathbf{S})$, it is necessary to obtain an estimate of $p(\mathbf{S})$, where all energetically relevant regions are ergodically sampled. Achieving an ergodic sampling of relevant configurations is a well-known problem that several enhanced sampling techniques have addressed.

For the ergodic sampling of a well-defined configuration space, metadynamics, which involves depositing penalty biases dynamically as the simulation proceeds to promote sampling,[16] would be a typical approach. However, for the exhaustive sampling of a conformation space, this approach is limited by the computational feasibility of storing the bias values on a grid of the same dimensionality as the conformation space, which is the same issue faced with the conventional method of FES construction. For this reason, in practical applications, conventional metadynamics biases constructed in dimensionalities higher than three are very rare. Alternatives for high-dimensional CV spaces, such as bias-exchange metadynamics,[17] have been developed. Still, these generally depend on the running of multiple replica simulations overseen by an exchange scheme. To widely sample conformational space with a single simulation, concurrent[18] well-tempered[19] metadynamics (WTMetaD) is used here.

Figure 2a shows a FES for alanine dipeptide computed from a nonconcurrent WTMetaD simulation using what will be referred to as the conventional method.[20] The WTMetaD biases are deposited in the 2D conformational space, defined by two torsions, ϕ and ψ (as illustrated in Figure 2d), ensuring that the entire space is fully sampled over the course of the simulation. The space is split into a 100 × 100 bins histogram. The distribution of MD configurations throughout the histogram follows the system's equilibrium probability distribution as distorted by the metadynamics bias. The total bias deposited in each bin is known, allowing this distortion of the probability distribution to be reweighted. The resulting FES has a resolution equal to the fineness of the grid, in this case $(2\pi)/100$ rad. This methodology is robust and widely adopted, but scales poorly to higher-dimensional FESes. Both the bias deposition during the simulation and the estimate of the probability distribution require the construction of a grid with the same dimensionality as the conformation space. If the same resolution is desired, increasing the number of torsions incurs an exponential cost on computational resources, rapidly becoming unfeasible. In the scientific literature, this issue is addressed by employing dimensionality reduction methods such as Sketch-map.[21,22] In these methods, the conformational probability density is obtained by histogramming sampled configurations in a low-dimensional space of unphysical coordinates. While this approach can be effective for relatively small systems,[22] its ability to resolve degeneracies and identify *conformers* when the conformational spaces are defined by a large number of torsional degrees of freedom is not straightforward.

Biasing and Estimating Probabilities in High Dimensions. *Gridless Probabilities with Density Peaks Advanced.* Density-based clustering techniques form a family of unsupervised machine learning algorithms that group data points within spatial data sets into clusters based on the distance between data points within the data space. Algorithms in this family include DBSCAN[23] and Fast Search and Find of Density Peaks (FSFDP).[24] There is precedent for the use of FSFDP in molecular conformation spaces, Marinova et al. used it to study the conformation space of Sildenafil.[6] Here, Density Peaks Advanced (DPA),[25] a successor to FSFDP, is used.

DPA, developed by d'Errico et al., splits a set of data points distributed in space into clusters by grouping points within density peaks, a term referring to regions of high data density (N.B. In this work, the term "density" refers to the density of data points in $\mathbf{S}$, unless otherwise noted). It does this partly by calculating the local density of every region centered on every single point in the data set. This calculation is a function of the Euclidean distances between the point and its nearest neighbors. Here, the process is applied to a sample of N configurations in conformation space sampled by the MD simulation. These local density calculations are extremely powerful in this context for two key reasons: first, each additional dimension in $\mathbf{S}$ adds a

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

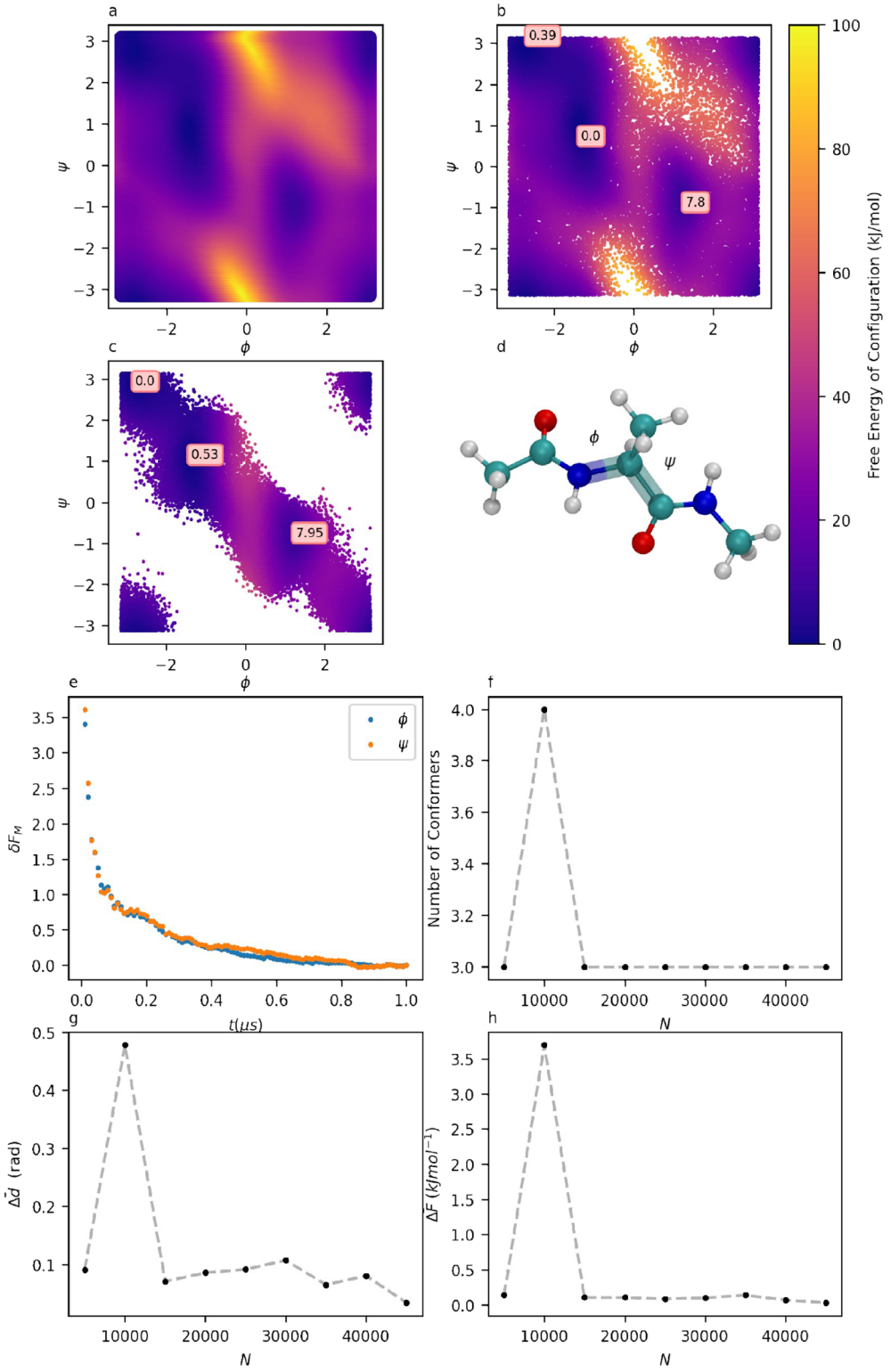


Figure 2. a: A conformational free energy surface for alanine dipeptide, obtained conventionally, through constructing a reweighted probability distribution on a histogram. b: The same free energy surface, constructed from a probability distribution derived from the local densities of individual configurations sampled from the simulation. The positions and free energies of local minima, as identified by DPA, are overlaid. c: The same free energy surface, also constructed from local densities, sampling a simulation using 2 × 1D WTmetaD biases instead of a conventional 2D bias. The positions and free energies of local minima, as identified by DPA, are overlaid. d: Alanine dipeptide, with the two relevant torsions ϕ and ψ highlighted. For a, b, and c, the energies of the FESes are indicated by a colormap in units of kJ mol^{-1}. e: Evolution of $\delta F(t)$ on the marginal free energies of each torsion in alanine dipeptide. f: Number of clusters identified by clustering on data sets of size N. g: Evolution of $\Delta\bar{d}$ with N for alanine dipeptide. h: Evolution of $\Delta\bar{F}$ with N for alanine dipeptide.

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

single term to the Euclidean distance calculation, so the cost with increasing dimensionality scales linearly, and second these local densities map a distribution in much the same way as the previously described histogram, so the same reweighing and inversion procedure may be applied to calculate the free energy. These free energies, unlike those in the above histogram, are not associated with a defined region of conformation space. Rather, they are associated with a specific configuration sampled by the simulation. Thus, this *per point* FES has no fixed spatial resolution; data is rich in regions that have been heavily sampled and sparse in regions that have not been frequented. This is advantageous as it means that while data in the local minima remains highly dense due to the frequent sampling, little cost is incurred in considering data from the rarely visited, largely irrelevant high-energy regions. This contrasts with the grid-based approach, where these high-energy regions are modeled in as high a resolution as the more relevant local minima. The results of applying this approach to alanine dipeptide can be seen in Figure 2b.

Biasing in High-Dimensional Spaces with Concurrent Metadynamics. To avoid the exponentially increasing costs of depositing WTMetaD biases in a high-dimensional conformational space, concurrent metadynamics is used in its place to promote sampling. This entails simultaneously depositing a single one-dimensional bias for each torsion in the molecule, thus encouraging exploration of the rotation of that torsion.[16] The cost of this approach scales linearly with dimensionality (shown in Figure S1a), as one additional monodimensional grid is needed for each additional torsion considered. The cost savings of this approach come with a trade-off; conventional metadynamics promotes the exploration of the entire conformational phase space and guarantees that previously visited configurations will be penalized accordingly. Concurrent metadynamics does not explicitly bias the combinations of any torsion values. Instead, it promotes the escaping from local free energy wells by driving the rotation of individual torsions. This can be seen in the use of this technique on alanine dipeptide in Figure 2c.

This work will demonstrate that DPA analysis of data sets generated using concurrent WTmetaD can be used to model the probability distribution of a flexible molecule's conformational state and create a "per-point" FES, where the data points themselves are configurations sampled by the simulation. This approach will be shown to recreate the well-studied 2D FES of alanine dipeptide before being demonstrated on 4- and 11-dimensional conformational free energy landscapes in vacuum and a 7-dimensional conformational free energy landscape in different solution environments.

THEORETICAL BACKGROUND AND METHODS

Clustering and FES Construction. To discuss the approach developed in this work in detail, it is helpful to begin by introducing how DPA allows the estimation of a FES from unbiased MD-generated data. Density Peaks Advanced (DPA) is run on a subset of configurations sampled from an MD trajectory. The size of this subset is the limiting factor in the cost of this approach, as each point has a local density determined by the distance to its neighbors, so the process depends on constructing a complete distance matrix between all configuration pairs. We, therefore, expect the cost of this approach to scale approximately with the square of the data set size (see Figure S1b). More advanced implementations that compute distances of k neighbors can achieve scaling closer to $N \log N$, with N being the size of the data set.[26] Because the thermodynamics of the system directs the MD trajectory's sampling, the local density of each configuration is proportional to the relative probability of encountering this configuration. It thus can be directly inverted to the free energy of the configuration using eq 1.

DPA estimates the probability associated with the ensemble of configurations projected in a given point i of the configuration space $\mathbf{S}$, using the PAk density estimator,[27] based on the Euclidean distances between point i and its k nearest neighbors. An underpinning assumption of this method is that the density is constant in the neighborhood of the point i. Hence, the parameter k is selected to be as large as possible to maximize the data used in calculating the local density while still representing a hypervolume of constant density. Each neighbor l can be said to occupy the volume v_l of the hyperspherical shell enclosed between hyperspheres of radii r_l and r_{l-1}. The sum of these volumes up to neighbor k is equal to the volume V_k of a hypersphere with radius r_k. DPA leverages the fact that for a region of constant density, the volumes will be drawn from an exponential distribution[28] with a rate of this density ρ and that thus the log-likelihood function of ρ given a set of k neighbors is provided by

$$L_{i,k}(\rho) = k \log(\rho) - \rho V_k \tag{2}$$

The PAk estimator selects an appropriate k value using two models with distinct assumptions. Model one, $M1$, assumes that the densities of point i and its $j = k + 1$ nearest neighbor are independent, while model two, $M2$, assumes these densities are identical. Their log-likelihood functions are

$$L_{M1} = k \log\left(\frac{k^2}{V_k V_j}\right) - 2k \tag{3}$$

$$L_{M2} = 2k \log\left(\frac{2k}{V_k + V_j}\right) - 2k \tag{4}$$

The two models are compared with a likelihood ratio test,[29]

$$D_k = -2(L_{M2} - L_{M1}) \tag{5}$$

which increases as the two models differ. If D_k grows over a threshold D_{thr} ($D_{\text{thr}} = 23.928$ according to ref 25) then the densities of i and j cannot be considered constant. As such, PAk selects an appropriate k value by iteratively calculating D_k for increasing values of k until the threshold is passed.

Per Point Free Energies with DPA and Biased Simulations. In practice, we generate conformational data sets using WTmetaD simulations. Using PAk on configurations sampled from a WTmetaD simulation produces densities that reflect a probability distribution perturbed by the applied biases. These densities can be reweighed using the Zwanzig approach[15] as

$$\rho_i^* = \rho_i \times e^{\beta(\sum_{t=1}^{D} V_i^t)} \tag{6}$$

where ρ^* is the reweighed density, β is equal to $1/k_B T$, V_i^t is the bias in torsion t, $\sum_{t=1}^{D} V_i^t$ represents the sum of the concurrent biases acting on the D torsions, for configuration i. Practically, we evaluate ρ, the biased density, from configurations generated in a quasi-static bias regime, as the bulk of the bias is deposited during the early stages of the simulation, and the bulk of sampled configurations are visited when bias deposition is negligible. We

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

therefore apply the *final bias* approximation to obtain a time-independent value of V_i^t acting on configuration i.[30−32] To mitigate the noise introduced by exponential reweighting,[15] the density of each point is then averaged over hyperspherical domains of radius 0.1 rad. This step generates a new smoothed set of densities $\overline{\rho}_i^*$, at the cost of a slight controllable loss in spatial resolution. The free energy F_i, associated with configuration i (thus termed *per point*), is computed as $F_i = -k_B T \ln \overline{\rho}_i^*$.

Conformer Classification. In the classification step, DPA identifies peaks in the density as cluster centers, i.e., distinct conformers. This operation is equivalent to identifying local minima in the D-dimensional free energy surface. For this step, we use the set of reweighted, smoothed densities $\overline{\rho}_i^*$.

Moreover, to avoid every fluctuation in density from being identified as a distinct peak, the DPA classifier is set to merge clusters separated by a saddle point between the free energy basins (a conformational transition state) with free energy less than 1 kT higher than one of the cluster centers it connects. Once cluster centers have been determined, all remaining configurations are assigned membership to the same cluster as their nearest neighbor of higher density.[25] With all configurations classified, clusters with a population smaller than 1% of the total sample are discarded to avoid spurious clusters identified from anomalously isolated configurations.

The ultimate product of this process is a set of cluster center configurations representing the local minima of the D-dimensional conformational FES, called a cluster set. Each cluster's lowest free energy configuration, i.e., the cluster center, provides *the most* representative configuration of a given conformer.

Consistency Analysis. The potential high dimensionality of $\mathbf{S}$ makes the visualization of per-point free energies difficult. A series of checks on the ergodicity of the sampling and consistency of the DPA classification, therefore, provides confidence in the results.

First, the convergence of the D monodimensional marginal free energy surfaces of each torsion is monitored to assess the ergodicity of the sampling. For a simulation of length τ, convergence of the marginals is assessed by monitoring, on D histogams of n_{hist} points, the quantity:

$$\delta F_M(t) = n_{\text{hist}}^{-1} \sum_{i=1}^{n_{\text{hist}}} |F_i^\tau - F_i^t| \tag{7}$$

Where simulation time t runs from $[0,\tau]$, $F(t)$ is a monodimensional FES obtained with data gathered up to time t, $F(\tau)$ is the same quantity computed with all the data available. This quantity represents the average free energy difference per histogram bin in any of the D monodimensional free energy surfaces.

Figure 2e displays an example of $\delta F_M(t)$ computed for ϕ and ψ torsional angles of alanine dipeptide during concurrent metadynamics. The flattening of these differences as the fraction of utilized trajectory increases indicates that the simulation has been run for a sufficiently long time, allowing these torsions to be ergodically sampled.

This check is computationally inexpensive and provides a first qualitative assessment of the quality of the configurational exploration obtained with concurrent WTmetaD. If the marginal FES associated with a torsion is still evolving rapidly at time τ, i.e., when the simulation ends, the sampling has not yet reached the ergodic limit with respect to the configurations discovered.

However, the convergence of 1D marginal FESes tells us little about exploring the conformational space in its full dimensionality. This is important as even substantial amounts of data may be distributed extremely sparsely in high dimensions.

As such, to build confidence in our results, we evaluate the statistical significance of the conformer classification as a function of the data set size.

For this purpose, a consistency check has been devised, which offers a similarity score between two cluster sets generated from different configurations. A cluster set generated from a data set of size N, C^N can be compared with a reference cluster set C^{ref}, which is generated with the largest number of configurations feasible. Each cluster center C_i^N is matched with the nearest center in the reference set C_i^{ref}, according to the Euclidean distances in $\mathbf{S}$ between members of the two cluster sets. This matching process is demonstrated in Figure 1. Differences in free energy ΔF_i and position Δd_i are determined and averaged across all matched pairs as $\Delta\overline{F}^N$ and $\Delta\overline{d}^N$. This comparison to C^{ref} can be repeated for cluster-sets generated from data sets of increasing N, and evolution of $\Delta\overline{F}^N$ and $\Delta\overline{d}^N$ with growing N can thus be assessed. Once data sets are large enough, the positions and relative free energies of minima would be expected to be independent of data set size. The results of this analysis on the case of alanine dipeptide are shown in Figure 2f,g,h.

Simulation Setup. Unless otherwise specified, all simulations carried out for this work consisted of a single molecule in vacuum, simulated with a 2 fs time step. GAFF[33] force field parameters were used, and GROMACS[34] was the MD engine used. WTMetaD was carried out using the Plumed[35] plugin for GROMACS. The simulations were carried out in the NVT ensemble, at a temperature of 300 K, maintained using the velocity-rescaling thermostat developed by Bussi et al.[36]

For the determination of WTMetaD parameters, a short 10 ns unbiased simulation was run. The marginal FES in each torsion was computed. A Gaussian Mixture Model was fitted to the resulting FES, and the smallest width parameter of the GMM, corresponding to the narrowest local minimum in the marginal FES, was considered as the minimum reference width for the marginal under consideration. The width of the Gaussian terms used to update the metadynamics bias was set to a quarter of the minimum reference width. Following the completion of the WTMetaD simulation, configurations from the simulation were paired with the total deposited bias at the corresponding position in conformation space, which aligned with the final bias approximation.

RESULTS AND DISCUSSION

Method Validation: Alanine Dipeptide. The workflow outlined in this work was first tested on the Ramachandran plot[37] of alanine dipeptide. This system was chosen for several reasons: the Ramachandran plot of alanine dipeptide is a commonly used model system in the field of molecular dynamics and enhanced sampling techniques, making it one of the best-studied conformational FESes available. Additionally, its low dimensionality allows for both the visualization of the FES and access to more conventional methods of exploring this conformational space. The principal results of this are shown in Figure 2. The structure of alanine dipeptide, with ϕ and ψ, indicated, is shown in Figure 2d. The conventional FES of alanine dipeptide, obtained through histogramming and

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

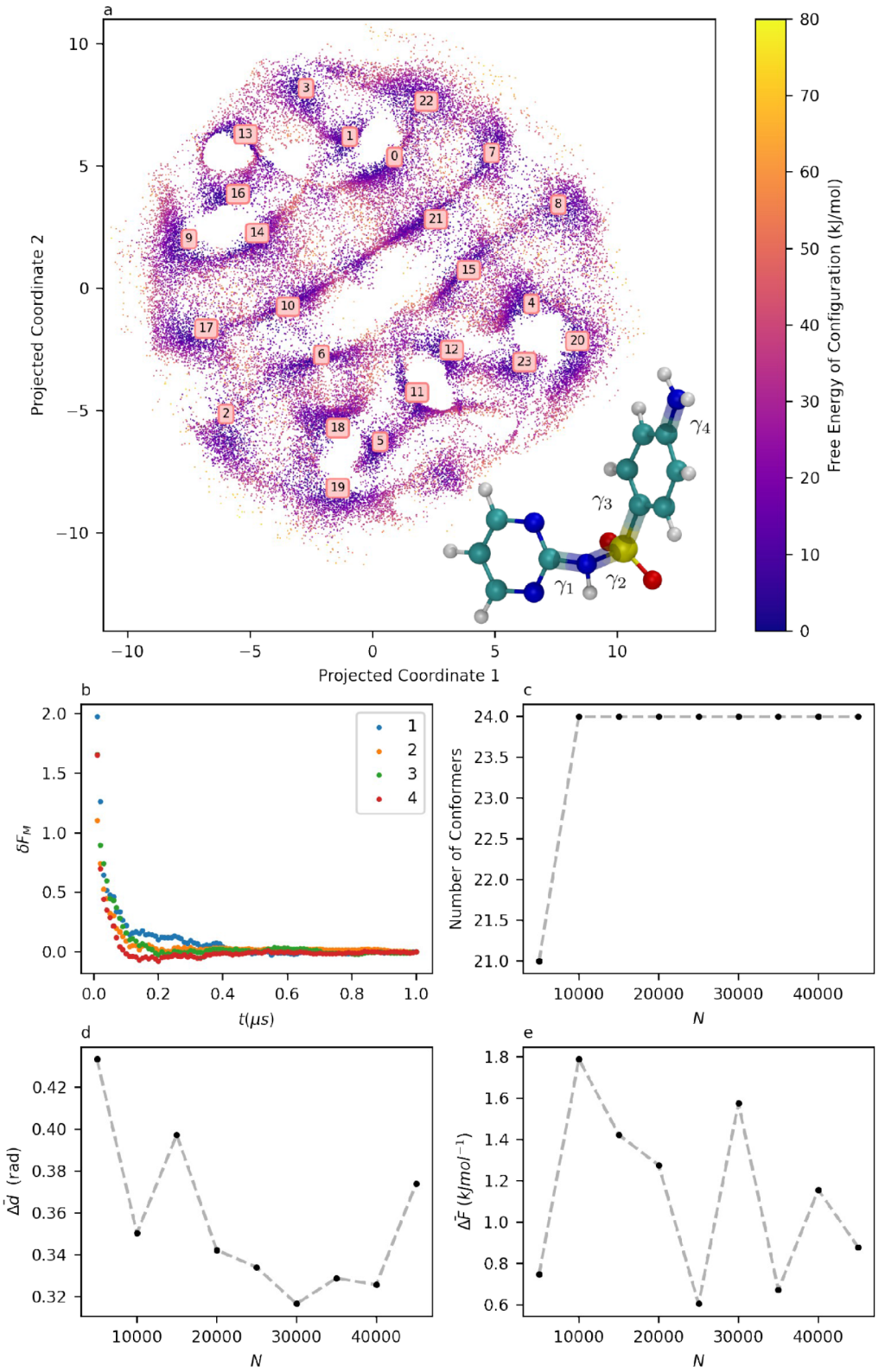


Figure 3. a: 2D Sketch-Map projection of sulfadiazine's 4D conformation surface, with molecular structure of SDZ inset. Distances between configurations are preserved over small separations, but the axes themselves have no physical meaning. b: Evolution of $\delta F(t)$ on the marginal free energies of each torsion in sulfadiazine. c: Number of clusters identified by clustering on data sets of size N. d: Evolution of $\Delta\bar{d}$ with N for sulfadiazine. e: Evolution of $\Delta\bar{F}$ with N for sulfadiazine.

reweighing of a trajectory generated through a WTMetaD MD simulation,[20] is shown in Figure 2a. A per-point FES, generated from the same simulation but with free energies calculated using the local densities of the sampled configurations, as discussed in the Methods section, is shown in Figure 2b. From a visual comparison, it clearly appears that the two FESes are in

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

agreement, demonstrating that the reweighted-DPA density estimate leads to results virtually indistinguishable from standard histogramming-based approaches. Figure 2c shows a per-point FES generated using a trajectory from a concurrent metadynamics simulation where ϕ and ψ are biased independently. Besides demonstrating the consistency of the free energies obtained by concurrent biasing, the comparison between Figure 2b and Figure 2c illustrates the trade-offs entailed by using concurrent metadynamics. As detailed in the Methods section, concurrent metadynamics promotes the sampling of metastable states without guaranteeing an exhaustive sampling of the joint configurational probability density. Nevertheless, all relevant free energy minima are adequately sampled, and their positions and free energies agree with those obtained by standard, two-dimensional metadynamics (Figure 2).

The results of the consistency analysis techniques outlined in the Methods section on the per-point FES outlined in Figure 2c are shown on Figure 2e−h. Figure 2e shows the evolution of $\delta F(t)$ for ϕ and ψ. The flattening of the curves shows that the sampling of the two torsions is indeed ergodic over the time scale of the simulation.

Figure 2h shows mean conformer free energy difference $\Delta\overline{F}^N$ (defined in the Methods section) obtained from clustering data sets of increasing N and a reference data set at N = 50,000 configurations. Figure 2g shows a similar plot displaying the mean separation of the cluster centers, $\Delta\overline{d}$. Figure 2f shows the number of minima identified by DPA for each reduced-size data set. The plot shows that all reduced data sets agreed that there were 3 conformers, with the exception of the 10,000 configuration data set. In all other data sets, there is very good agreement on the position and free energies of the local minima, with energy differences well within 1 kJ mol^{-1} and mean separations hovering around 0.1 rad. It is surprising how little data is required to generate reasonable results in this 2-dimensional case. Even in Figure 1, where very small data sets are used to illustrate the cluster-set matching procedure, the agreement in minima positions is apparent.

Applications to Higher-Dimensional Free Energy Surfaces. Having demonstrated the workflow developed here on the two-dimensional case of alanine dipeptide , higher-dimensional cases are now explored, where visualization of the entire conformation space is not possible, and conventional grid-based methods become unfeasible. Sulfadiazine, with a 4-dimensional conformational space, and Candidate XXXII, from the CSP Blind Test,[8,38] with an 11-dimensional conformational space, serve as a test for the ability of the workflow to handle conformational complexity. Sketch-map[21] is used in these cases to project a 2D representation of the high-dimensional per-point FES for human interpretation.

Sulfadiazine. Sulfadiazine is an antibiotic molecule with a 4-dimensional conformational space; its clinical relevance and intermediate complexity make it an ideal next step for the method outlined here. A 4-dimensional space is too high to allow a FES to be fully visualized while still being low enough that reasonable data density can be obtained (50,000 data points in a periodic 4D space results in an average density of roughly 32 configurations per rad^4). The inset in Figure 3a shows the 4 torsions considered in sulfadiazine. Using the same approach outlined above for alanine dipeptide, sulfadiazine's conformational FES was studied by analyzing the configurations sampled within a 1 μs single-molecule WTmetaD simulation. The resulting per-point FES cannot be fully visualized without dimensionality reduction, so the relative free energies and coordinates of each minimum are presented in Table 1. Figure

Table 1. Labels, Free Energies, and CV-Space Coordinates of Sulfadiazine's 24 Conformers[a]

Conformer	Free energy [kJ/mol]	γ_1	γ_2	γ_3	γ_4
0	0.72	−1.71	−2.1	−1.51	2.95
1	1.28	−1.7	0.13	−1.52	−2.98
2	0.75	1.55	−1.83	1.54	0.21
3	1.44	−1.62	−0.13	−1.47	−0.05
4	0.08	−1.81	−1.94	1.62	−2.93
5	1.05	1.64	0.09	1.58	2.86
6	0.0	1.92	1.75	1.42	−0.28
7	1.03	−1.43	1.93	−1.31	0.07
8	1.66	−1.49	2.08	1.26	0.23
9	0.67	1.3	−1.87	−1.6	−3.02
10	0.63	1.68	2.05	−1.66	0.15
11	1.56	1.62	−0.11	1.68	0.21
12	1.81	−1.55	0.01	1.64	0.17
13	0.21	1.62	−0.16	−1.65	−0.15
14	0.93	1.85	1.89	−1.53	3.13
15	1.02	−1.76	−1.95	1.88	0.31
16	1.2	1.62	0.06	−1.7	−3.08
17	1.4	1.4	−1.96	−1.77	0.15
18	1.58	1.74	1.88	1.63	−3.07
19	0.21	1.55	−2.04	1.83	2.93
20	0.8	−1.55	2.0	1.54	−2.81
21	0.47	−1.82	−1.83	−1.59	0.2
22	1.76	−1.46	2.0	−1.77	2.92
23	1.03	−1.58	0.01	1.69	2.9

[a]The labeling convention is consistent with that of Figure 3.

3a shows a 2D projection of the 4D per-point FES created using SketchMap. This representation preserves the short-distance connectivity between data points, allowing for the visualization of distinct free energy basins and the transition states between them, although the two axes of the new 2D projection are not physically meaningful in themselves.[21] It should be emphasized that the estimation of densities and the determination of the number and coordinates of the free energy minima are determined in the full 4-dimensional conformation space and that the projection in Figure 3a serves only to assist in the visualization of the relationships between different conformers. It is possible to combine the 4-dimensional information presented in Table 1 with the 2-dimensional intuition provided by Figure 3a. For example, the FES in Figure 3a appears to be bisected by a diagonal channel, and indeed, by inspecting the torsion values of the conformer pairs 17 and 2, 10 and 6, 21 and 15, and 7 and 8, it can be determined that these conformers pairs are equivalent, and differ from each other in a symmetric rotation of π radians of γ_3. This example illustrates how these 2D projections can be interpreted and demonstrates how symmetry elements in the molecule's conformational space can be preserved in the 2D projection.

The results of the consistency metrics for sulfadiazine are shown in Figure 3b−e. In comparing these results to those in Figure 2e−h, it is possible to evaluate the impact of doubling the dimensionality of the conformation space on the accuracy and data efficiency of the classification process. The plots of $\delta F(t)$ for the four torsions show that the four marginals in Figure 3b converge rapidly, providing evidence of ergodicity. Figure 3c shows that, except for the 5000-point data set, repeated analyses

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

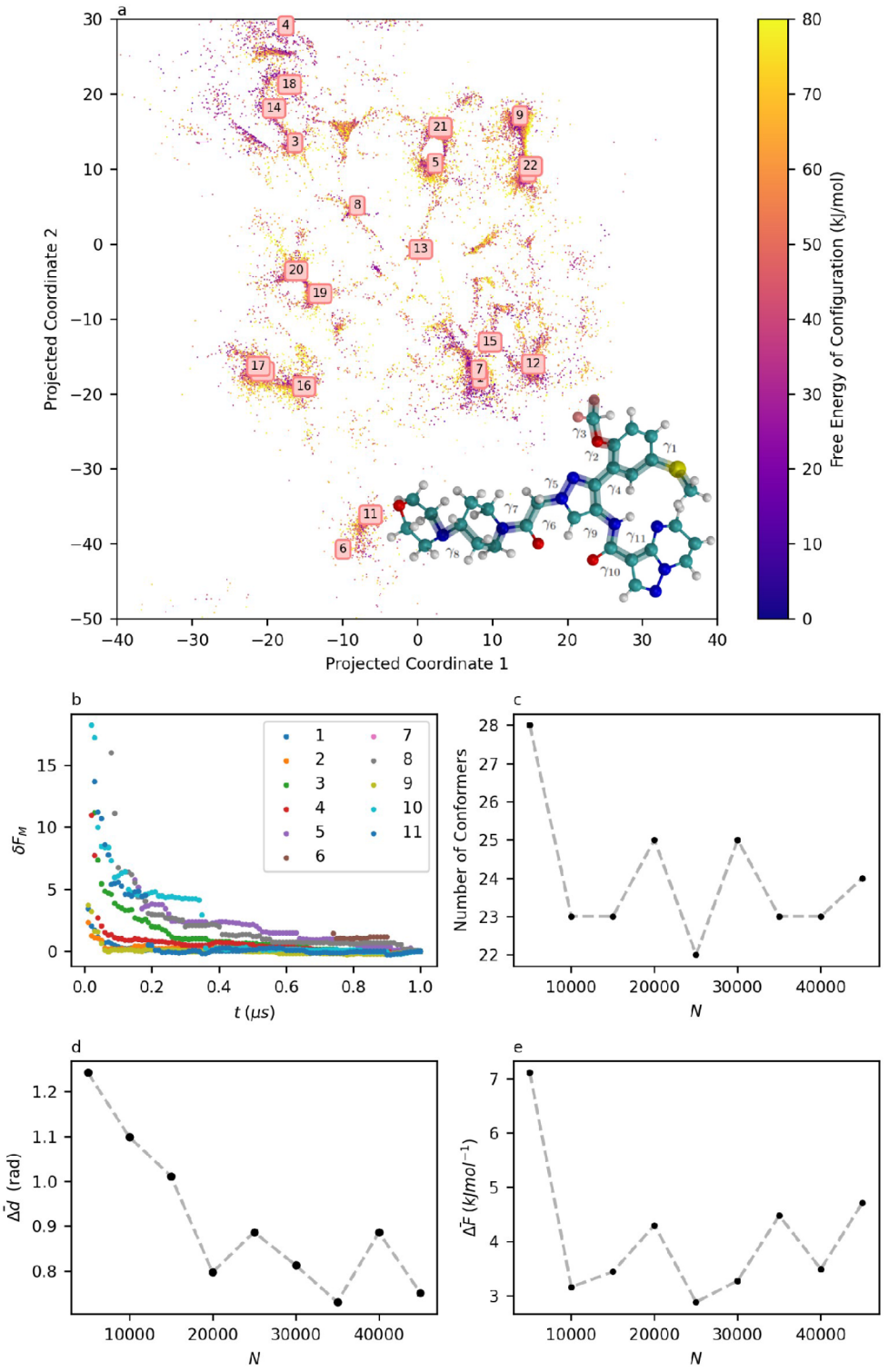


Figure 4. a: 2D Sketch-Map projection of XXXII's 11D conformation surface, with molecular structure of XXXII inset. Distances between configurations are preserved over small separations but the axes themselves have no physical meaning. b: Evolution of $\delta F(t)$ on the marginal free energies of each torsion in XXXII. c: Number of clusters identified by clustering on data sets of size N. d: Evolution of $\Delta\bar{d}$ with N for XXXII. e: Evolution of $\Delta\bar{F}$ with N for XXXII.

achieve a consistent number of 24 conformers. Figure 3d,e shows the evolution of $\Delta\bar{F}$ and $\Delta\bar{d}$ respectively, as N increases to a reference value of 50,000. Here, the differences between alanine dipeptide and sulfadiazine become apparent. The mean free energy deviation jumps from being nearly negligible to a range between 0.5 and 2 kJ mol^{-1}, and positional deviation increases from approximately 0.1 rad to between 0.3 and 0.4 rad. Sulfadiazine's energy deviation is still within 1 k_BT, and the positional deviations still correspond to very small changes in the molecular structure. However, the abrupt change following an increase in dimensionality highlights the importance of carrying out consistency checks when working with highly

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

Table 2. Labels, Free Energies, and CV-Space Coordinates of XXXII's 23 Conformers[a]

Conformer	Free energy [kJ/mol]	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	γ_{11}
0	6.09	1.12	−1.27	0.89	0.34	−2.31	1.13	0.42	1.29	2.96	3.13	0.15
1	0.0	3.06	1.84	−1.52	−2.75	−2.24	1.29	0.2	1.12	2.95	−3.13	0.04
2	8.45	1.03	1.98	1.07	0.37	−2.24	1.12	0.85	−1.19	−3.05	3.09	0.12
3	6.2	1.25	1.24	1.05	−2.93	2.43	−1.11	2.65	−1.12	3.11	3.11	−0.08
4	3.4	1.21	−1.05	1.72	2.95	2.23	−1.21	2.28	1.1	−3.09	2.93	−0.1
5	8.69	1.17	1.71	0.85	0.58	−2.06	1.21	−0.22	1.19	−3.06	−2.94	0.2
6	15.67	3.04	−0.8	−1.39	−0.41	1.06	0.93	0.43	1.42	3.07	3.13	−0.34
7	2.17	−3.06	1.14	−1.6	−2.85	−2.0	1.39	−0.18	1.0	3.1	−2.94	0.09
8	16.12	1.07	−0.62	1.61	2.67	0.91	1.01	0.26	0.94	−3.04	3.02	−0.3
9	7.62	1.03	−1.75	0.93	0.56	−2.35	1.25	0.69	−1.06	2.97	−3.08	0.18
10	12.45	−3.09	−0.61	−1.43	−0.3	2.28	−1.13	2.41	1.11	2.99	3.09	−0.18
11	21.03	2.81	1.0	−1.16	−0.36	1.04	1.0	0.51	1.14	2.92	−3.02	−0.31
12	1.43	−3.1	1.39	−1.58	−2.64	−2.34	1.17	0.64	−0.89	−3.03	3.12	0.07
13	6.92	−2.99	1.63	−1.6	−2.7	−1.02	−0.97	2.73	−0.93	2.83	−2.92	0.04
14	5.96	1.17	−0.86	1.51	2.86	1.96	−0.94	2.64	−1.04	3.12	−3.05	−0.09
15	5.72	3.04	0.71	−1.49	−2.51	−2.39	1.24	0.53	3.03	3.03	−2.98	−0.09
16	10.98	−3.01	2.07	−0.91	−0.57	2.26	−1.36	2.17	1.23	3.13	3.09	−0.16
17	8.91	−3.1	−1.4	−0.9	−0.42	2.12	−1.16	2.32	1.26	3.03	−2.94	−0.12
18	6.88	1.18	−1.62	1.61	2.66	2.1	−1.19	2.85	−0.95	−3.12	2.92	0.03
19	9.53	3.02	1.55	−1.25	−0.56	2.09	−1.2	2.69	−0.91	2.98	3.11	−0.09
20	6.89	3.07	−2.12	−1.31	−0.62	2.07	−1.17	2.86	−1.06	3.06	3.11	−0.21
21	7.35	1.18	−2.3	1.2	0.55	−2.11	0.89	0.41	1.37	2.86	−2.81	0.09
22	6.13	1.16	1.23	0.85	0.27	−2.15	1.19	0.59	−0.89	3.14	−2.93	0.16

[a]The labeling convention is consistent with that of Figure 3.

unintuitive results that are difficult to inspect visually. Due to the number of equivalent conformers related to one another by symmetric transformations in sulfadiazine, it is possible to compare the free energies of equivalent conformers as an assessment of the reproducibility of the free energy calculation. This is not recommended as a standard practice, as the presence of symmetrically related conformers is system-dependent and not guaranteed. However, in this case, comparing the differences between equivalent conformers reveals deviations of the same order as the mean free energy deviations calculated in the smaller data sets (Figure 3e).

Target XXXII of the Seventh CCDC Blind Test. The highest-dimensional conformational FES explored here is that of Molecule XXXII, a target from the seventh CSP Blind Test.[8,38] As a highly flexible drug-like molecule with a conformation space defined by 11 torsions (shown in the inset in Figure 4a), it is chosen here to test the limits of our method. To facilitate comparison with results collected for alanine dipeptide and sulfadiazine, the results presented here were generated using consistent simulation and analysis parameters. Using 50,000 data points in this high-dimensional space results in an average data density of approximately 8×10^{-5} configurations per rad^{11}. Despite the extremely low data density, which is inherently linked to the complexity of the conformational space, we show that meaningful results are achievable. Figure 4a shows the projected 11-dimensional per-point FES, with cluster centers corresponding to 11-dimensional coordinates presented in Table 2. When comparing this FES to sulfadiazine's in Figure 3a, the features of XXXII can be seen reflected in its own FES. The relative lack of symmetrical torsions results in a less symmetrical FES, and the higher-dimensional FES is much sparser, illustrating that the computational savings arise from a more efficient, rather than more exhaustive, sampling of conformational space.

The consistency metrics in Figure 4b–e are, however, less reliable than those obtained for sulfadiazine. Figure 4b shows well-converged marginal free energies, but Figure 4c shows that the number of conformers identified is less consistent than in lower-dimensional cases. The number of metastable states identified as distinct conformers hovers between 22 and 25 for data sets sized 10000 and upward. Along with a fluctuating number of conformers, larger deviations in free energies and positions are now observed, with $\Delta\overline{F}$ between cluster sets now varying by up to 5 kJ mol^{-1}, and $\Delta\overline{d}$ drifting by as much one full radian, even at large data set sizes. Despite this drop in the quality of the results, we believe it is still remarkable that a reasonably intuitive understanding of such a high-dimensional conformational FES can be derived from a limited amount of data in a computationally accessible way, even if its value in this instance is chiefly qualitative. To further explore the consistency of the FES in Figures 4, and S2–S10 contain the FES projection for each of the smaller data sets used in the consistency analysis, allowing the evolution of this per-point FES to be observed. Inspection of this evolution in the FES appears to reveal that the majority of fluctuations in $\Delta\overline{F}$ and $\Delta\overline{d}$ observed arise in higher-energy conformers, with the low-energy regions converging at lower N values. Although we do not rigorously prove this here, it is a reasonable expectation, as lower energy regions have a high data density, resulting in more accurate free energy estimates based on a greater amount of data.

Exploring the Impact of Solvent on the Conformational Landscapes of Bicalutamide. DPA operates on a set of molecular configurations defined purely by the values of the subject molecule's torsions. As such, the cost of the analysis is independent of the length, complexity, and level of theory of the simulations from which the configurational data set is produced. Generating a high-dimensional conformational free energy landscape for a molecule simulated in a solvent environment is, therefore, accessible. In this section, a study of the molecule

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

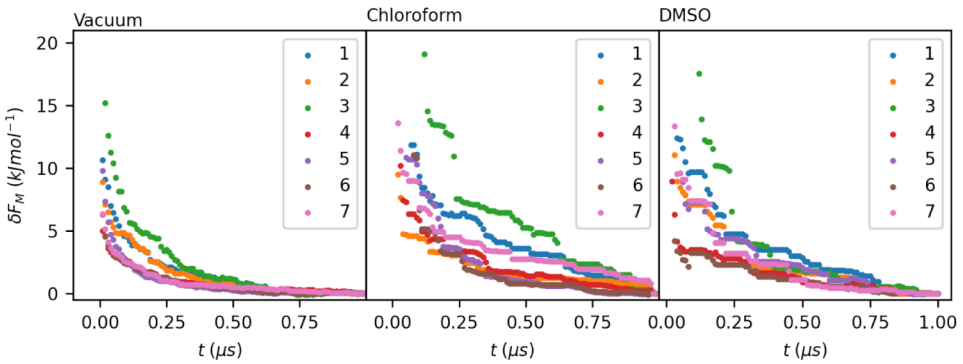


Figure 5. Evolution of $\delta F(t)$ on the marginal free energies of each torsion in bicalutamide, in vacuum, chloroform, and DMSO.

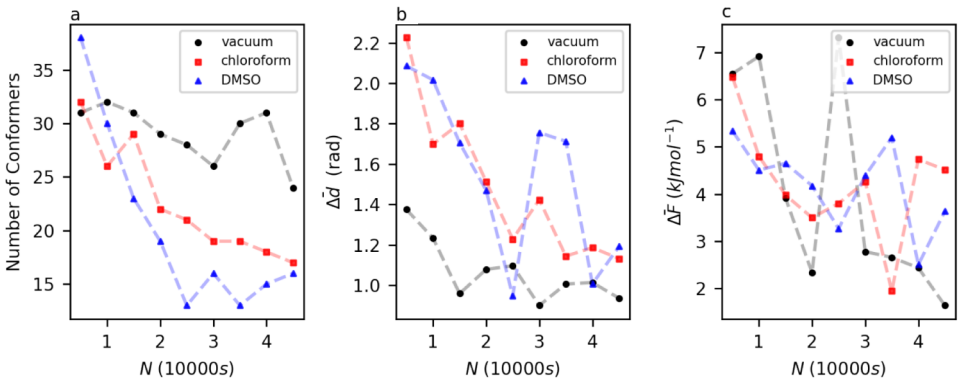


Figure 6. a: Number of clusters identified by clustering on data sets of size N for bicalutamide, in vacuum, chloroform, and DMSO. b: Evolution of $\Delta\bar{d}$ with N for bicalutamide, in vacuum, chloroform, and DMSO. c: Evolution of $\Delta\bar{F}$ with N for bicalutamide, in vacuum, chloroform, and DMSO.

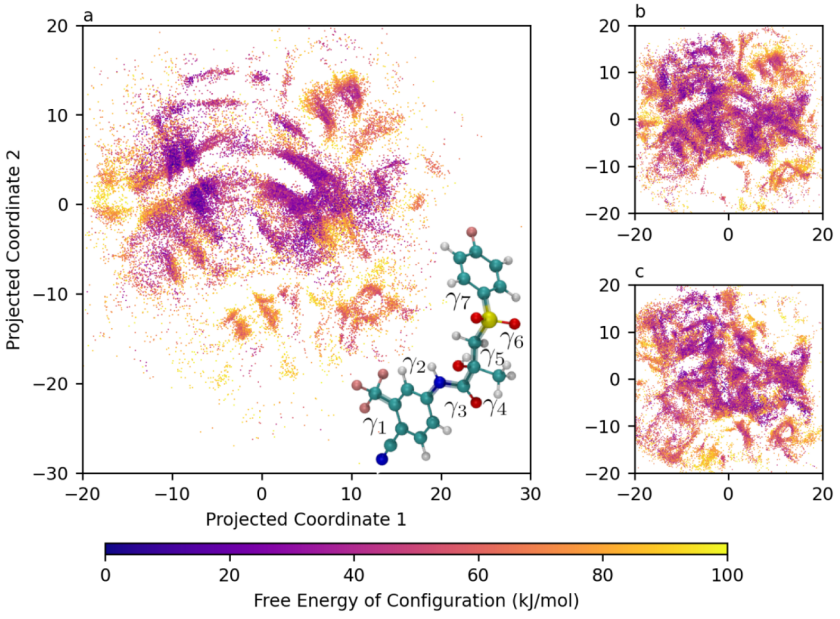


Figure 7. 2D Sketch-Map projection of bicalutamide's 7D conformational free energy landscape in vacuum (a), chloroform (b), and DMSO (c), with molecular structure of bicalutamide inset. Distances between configurations are preserved over small separations but the axes themselves have no physical meaning.

bicalutamide is presented, examining the impact of two different solvent environments on the conformational free-energy landscape.

Bicalutamide is an antiandrogen compound used for the treatment of prostate cancer. It is highly flexible, with a conformational space described by the 7 dihedral angles shown inset in Figure 7a. This flexibility results in two conformational polymorphs being observed in the solid state: form I, as shown in Figure 9c, and form II, as shown in Figure 9d. Form I demonstrates an open conformation, while form II adopts a more compact, closed conformation. Form I is the more thermodynamically stable form, and is the form which typically

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

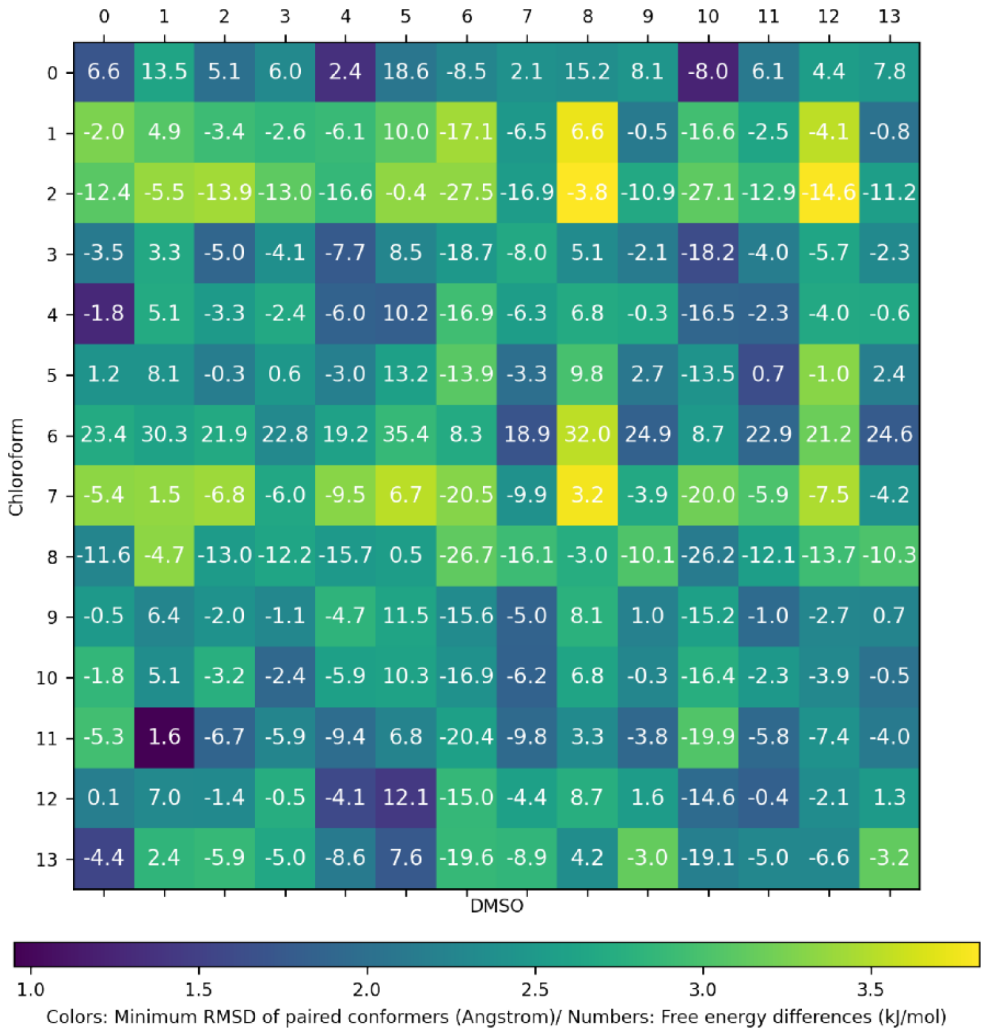


Figure 8. A dual matrix presenting a pairwise comparison of conformers of bicalutamide observed in chloroform (rows) and DMSO (columns). The color gradient indicates the minimum RMSD between atoms between the two conformers, while the number within each element indicates the stability of the conformer in chloroform relative to the conformer in DMSO.

arises upon recrystallization from most solvents.[39] Form II can be obtained from a melt of form I.[40] Despite form I's tendency to recrystallize out of most solvents, Sobornova et al. discovered that solvent choice had a significant impact on bicalutamide's conformational distribution in solution.[39] Using NOE (nuclear Overhauser effect) spectroscopy, they demonstrated that polar solvents promote an open conformation, while nonpolar solvents promote a closed conformation. Here, the conformational free energy landscapes of bicalutamide, simulated in vacuum, chloroform, and DMSO environments, are explored using the gridless method developed in this study. As before, data sets of 50,000 configurations were used to construct the per-point FESes shown here.

Figure 5 shows the evolution of $\delta F_m(t)$ for each of the 7 torsions of bicalutamide in vacuum, dichloromethane, and DMSO. This figure demonstrates the convergence of each of these one-dimensional marginal free energies. The results of the higher dimensional consistency checks are shown in Figure 6. Figure 6a shows that a consistent number of conformers was not reached in any of the environments simulated. This inconsistency in conformer number is characteristic of free energy landscapes computed in high dimensions, reflecting the trend seen with the consistency of Target XXXII. Despite the number of conformers continuing to fluctuate as the largest data set size is reached, the average positions of equivalent conformers are fairly similar, as demonstrated by Figure 6b. For the landscapes in vacuum and chloroform, equivalent conformers are found on average less than 1.2 Euclidean radians from each other in a 7-dimensional space. This number is slightly less consistent in DMSO, with some $\Delta\bar{d}$ values being as high as 1.8 Euclidean radians, even toward the final data set size. Finally, considering the difference in free energies between equivalent conformers, demonstrated in Figure 6c, it can be seen that in all three environments, conformers deemed equivalent are within, on average, 5 kJ mol^{-1} of each other. This is not as good an agreement as observed in the lower-dimensional cases; however, it will be sufficient when comparing the free energies of conformers that differ by more than this amount.

The per-point free energy landscapes generated through the gridless analysis are projected into two dimensions using Sketch-map. To enable comparison across different environments, the Sketch-map projections generated for the solvated cases use the same a, b, and σ parameters and landmark points as determined from the postprocessing of the simulation in vacuum. This ensures that the resulting two-dimensional projection is equivalent across the different environments. These projections

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

in vacuum, chloroform, and DMSO are shown in Figure 7a,b,c, respectively. For each of the projections, an enlarged image featuring the locations of the free energy minima marked with a numerical label is available in Figures S11−S13. These labels correspond to the conformer indices in the left-hand columns of Tables S1−S3 for bicalutamide in vacuum, chloroform, and DMSO, respectively. Some of the structure present in the vacuum projection seems to be preserved in the chloroform projection, with the left−right gulf having been narrowed slightly and conformers being distributed more diffusely. The DMSO projection, however, appears significantly different, with a new network of interconnected conformers being shown.

Figure 9 compares the conformational free energy minima in each solvent environment with the experimentally determined crystal structures. Figure 9a shows the most stable conformer in DMSO, which has an open conformation, as observed experimentally by Sobornova et al.[39] This conformation is distinct but similar to the conformation observed in form I of bicalutamide, illustrated in Figure 9c. The most stable conformer in chloroform is shown in Figure 9b and demonstrates a closed conformation, again matching the experimental observation of Sobornova et al.[39] Additionally, the conformation adopted in chloroform corresponds closely with the conformation adopted in crystal form II, shown in Figure 9d.

With confirmation that the most stable conformers in each environment align with experimental observations, a more extensive analysis of the relationships between distinct conformers and solvent environments can be carried out. Figure 8 shows a double matrix comparing all bicalutamide conformers in chloroform to all bicalutamide conformers in DMSO. Each element's number corresponds to the difference in free energy between the two conformers (relative to the most stable conformers in their environment). The color gradient indicates the similarity of the structures, measured by their minimum RMSD separation[41,42] considering all atom positions. In order to study these conformer sets more closely, we define the set of common conformers to be those conformer pairs that exhibit a minimum RMSD of less than 1.7 Å. There are 10 of these common conformers for bicalutamide in chloroform and DMSO, and they are shown in Figure 10. The overlapped conformers are shown in blue for chloroform and red for DMSO. It is interesting to note that the fully closed conformation observed as the most stable form in chloroform and in bicalutamide's crystalline form II, shown in Figure 9b,d, is not present as a common conformer, meaning it does not arise at all in DMSO. The majority of the common conformers shown in Figure 10 seem to exhibit a semiopen "L"-shaped conformation, rather than the fully open and closed conformations seen as the most stable conformers in DMSO and chloroform in Figure 9a,b. The conformers in Figure 10 are ordered by ΔF where

$$\Delta F = F_{\mathrm{CLF}} - F_{\mathrm{DMSO}}$$

where F_{CLF} and F_{DMSO} are the free energies of the conformers in chloroform and DMSO, respectively. Note that, as before, these individual free energies are themselves relative to the lowest energy configuration within the free energy landscape.

Figure 10 thus seems to show that these "L"-shaped conformers tend to be stabilized in chloroform, the same environment that promotes the fully closed conformer, and that common conformers with a greater open character tend to be stabilized by DMSO.

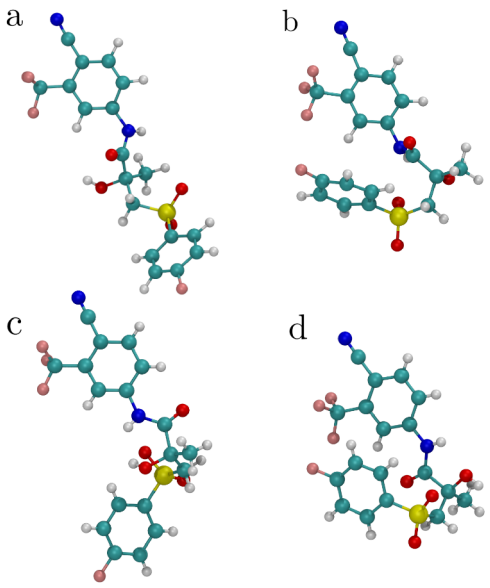


Figure 9. a: The lowest free energy conformer of bicalutamide in DMSO, index 5 Figure 7b. b: The lowest free energy conformer of bicalutamide in chloroform, index 2 in Figure 7c. c: The experimentally observed conformation of bicalutamide form I. d: The experimentally observed conformation of bicalutamide from II.

It is also interesting to consider the common conformer labeled CLF5-DMSO11, corresponding to conformer 5 in the chloroform landscape and 11 in the DMSO landscape. This conformation closely resembles bicalutamide crystal form I, as shown in Figure 9c. This common conformer has a low ΔF of 0.68 kJ mol^{-1}, indicating that it is equally stabilized by both solvents, but has an F_{CLF} of 16.22 kJ mol^{-1} and an F_{DMSO} of 15.54 kJ mol^{-1}, making it far from the most stable conformer in either solvent. From this, it can be inferred that while the form I conformation is metastable in solution, it is not purely a solvent effect that is responsible for the conformational rearrangements leading to the observed conformational polymorph.

CONCLUSIONS

We have introduced a gridless methodology for constructing high-dimensional conformational free energy landscapes from enhanced sampling simulations. By leveraging concurrent well-tempered metadynamics and Density Peaks Advanced (DPA) clustering, our approach enables the assignment of per-configuration free energies without resorting to dimensionality reduction or binning. This framework avoids the exponential cost of grid-based FES construction, making it particularly well-suited for flexible, drug-like molecules.

We validated the method across a range of molecular systems, from the benchmark alanine dipeptide to realistic pharmaceutical compounds with up to 11 torsions. Crucially, we demonstrated its application in explicit solvent environments, capturing solvent-induced shifts in conformational preferences of bicalutamide in quantitative agreement with experimental observations. The consistent performance across increasing dimensionality, together with the ability to capture solvent effects, underscores the robustness and transferability of the proposed approach.

Naturally, these features come with practical trade-offs. Generating sufficiently ergodic sampling for fully solvated systems requires molecular dynamics simulations on the scale of tens to hundreds of nanoseconds, which constrains the

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

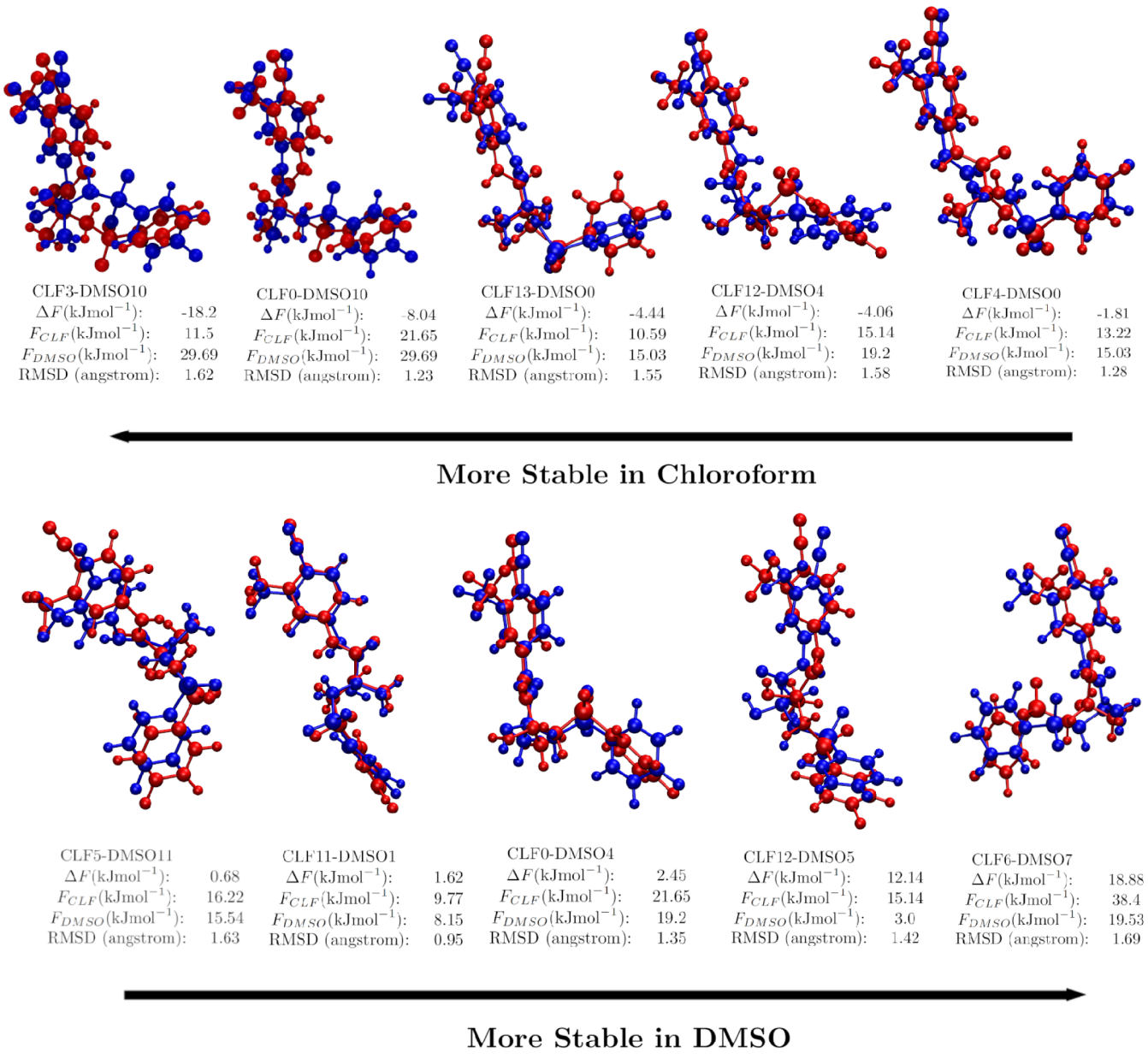


Figure 10. Ten common conformers of bicalutamide in chloroform and DMSO. For this system, conformers are deemed common to both solvents if their overlaid structures present a minimum RMSD deviation of less than 1.7 Å. Conformers in chloroform are indicated with the label CLF and have their molecular structures shown in blue. Conformers in DMSO are indicated with the label DMSO, and their molecular structures are shown in red. The free energy in both solvents, as well as the difference in free energy, is indicated for each conformer, and the conformers are ordered from those most stabilized in chloroform to those most stabilized in DMSO.

accuracy of the potential energy functions that can be feasibly employed. Moreover, because our approach relies on explicit sampling rather than energy minimization, it is inherently less suited for large-scale screening. While minimization-based workflows can be applied to thousands of compounds[9−12] our method is suited to rigorously investigating how explicitly represented environments affect APIs' conformational landscapes.

Because the method operates directly on torsional coordinates, it is agnostic to the underlying simulation engine or force field. As such, we envision its application in conjunction with machine-learned potentials as a key future development to complement extensive sampling with DFT-level accuracy. These features provide a practical and extensible tool for exploring conformational thermodynamics in atomistic simulations of complex molecular systems. The code developed here is fully open source and available from https://github.com/ucecvan/Twister.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c01247.

Supporting figures reporting sketchmap representations at different data set sizes, and tables summarizing free energy and position in dihedral angle space of bicalutamide conformers in various environments (PDF)

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480

AUTHOR INFORMATION

Corresponding Author

Matteo Salvalaglio − *Thomas Young Centre and Department of Chemical Engineering, University College London, London Wc1e 7je, United Kingdom;* orcid.org/0000-0003-3371-2090; Email: m.salvalaglio@ucl.ac.uk

Authors

Alexandre Ferreira − *Thomas Young Centre and Department of Chemical Engineering, University College London, London Wc1e 7je, United Kingdom*

Rui Guo − *Pfizer Worldwide Research & Development, Sandwich Ct13 9nd, United Kingdom*

Ivan Marziano − *Pfizer Worldwide Research & Development, Sandwich Ct13 9nd, United Kingdom;* Present Address: Particology Ltd., Discovery Park House, Ramsgate Road, Sandwich CT13 9ND, United Kingdom; orcid.org/0000-0002-3759-0070

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.5c01247

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge financial support by Pfizer and by the Engineering and Physical Sciences Research Council (EPSRC) (Grant No. EP/R018820/1). M.S. acknowledges funding from the Crystallization in the Real World EPSRC Programme Grant (EP/R018820/1) and from the ht-MATTER UKRI Frontier Research Guarantee Grant (EP/X033139/1). We are grateful to the UK Materials and Molecular Modelling Hub which is partially funded by EPSRC (EP/P020194/1) and the UCL High Performance Computing Facilities and associated support services for computational resources.

REFERENCES

(1) Cruz-Cabeza, A. J.; Bernstein, J. Conformational Polymorphism. *Chem. Rev.* **2014**, *114*, 2170−2191.

(2) Bernstein, J.; Hagler, A. T. Conformational polymorphism. The influence of crystal structure on molecular conformation. *J. Am. Chem. Soc.* **1978**, *100*, 673−681.

(3) Burcham, C. L.; Doherty, M. F.; Peters, B. G.; Price, S. L.; Salvalaglio, M.; Reutzel-Edens, S. M.; Price, L. S.; Addula, R. K. R.; Francia, N.; Khanna, V.; Zhao, Y. Pharmaceutical Digital Design: From Chemical Structure through Crystal Polymorph to Conceptual Crystallization Process. *Cryst. Growth Des.* **2024**, *24*, 5417−5438.

(4) Lucaioli, P.; Nauha, E.; Gimondi, I.; Price, L. S.; Guo, R.; Iuzzolino, L.; Singh, I.; Salvalaglio, M.; Price, S. L.; Blagden, N. Serendipitous isolation of a disappearing conformational polymorph of succinic acid challenges computational polymorph prediction. *CrystEngcomm* **2018**, *20*, 3971−3977.

(5) Marinova, V.; Wood, G. P. F.; Marziano, I.; Salvalaglio, M. Dynamics and Thermodynamics of Ibuprofen Conformational Isomerism at the Crystal/Solution Interface. *J. Chem. Theory Comput.* **2018**, *14*, 6484−6494.

(6) Marinova, V.; Dodd, L.; Lee, S.-J.; Wood, G. P. F.; Marziano, I.; Salvalaglio, M. Identifying Conformational Isomers of Organic Molecules in Solution via Unsupervised Clustering. *J. Chem. Inf. Model.* **2021**, *61*, 2263−2273.

(7) Elts, E.; Greiner, M.; Briesen, H. In Silico Prediction of Growth and Dissolution Rates for Organic Molecular Crystals: A Multiscale Approach. *Crystals* **2017**, *7*, 288.

(8) Hunnisett, L. M.; et al. The seventh blind test of crystal structure prediction: Structure generation methods. *Acta Crystallogr., Sect. B* **2024**, *80*, 517−547.

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

(10) Pracht, P.; Grimme, S.; Bannwarth, C.; Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.; et al. CREST—A program for the exploration of low-energy molecular chemical space. *J. Chem. Phys.* **2024**, *160*, 114110.

(11) Zurek, C.; Mallaev, R. A.; Paul, A. C.; van Staalduinen, N.; Pracht, P.; Ellerbrock, R.; Bannwarth, C. Tensor Train Optimization for Conformational Sampling of Organic Molecules. *J. Chem. Theory Comput.* **2025**, *21*, 1459−1475.

(12) de Souza, B. GOAT: A Global Optimization Algorithm for Molecules and Atomic Clusters. *Angew. Chem., Int. Ed.* **2025**, *64*, No. e202500393.

(13) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.

(14) Ferro-Costas, D.; Mosquera-Lois, I.; Fernández-Ramos, A. TorsiFlex: An automatic generator of torsional conformers. Application to the twenty proteinogenic amino acids. *J. Cheminf.* **2021**, *13*, 100.

(15) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

(16) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. rev.: Comput. Mol. Sci.* **2011**, *1*, 826−843.

(17) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.

(18) Gil-Ley, A.; Bussi, G. Correction to Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *J. Chem. Theory Comput.* **2015**, *11*, 5554.

(19) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(20) Shell, M. S.; Panagiotopoulos, A.; Pohorille, A. *In Free Energy Calculations*, Chipot, C.; Pohorille, A., Eds.; Springer, 2007, pp. 77−116.

(21) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13023−13028.

(22) Comitani, F.; Rossi, K.; Ceriotti, M.; Sanz, M. E.; Molteni, C. Mapping the conformational free energy of aspartic acid in the gas phase and in aqueous solution. *J. Chem. Phys.* **2017**, *146*, 145102.

(23) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 1996, *96*, 226−231.

(24) Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492−1496.

(25) d'Errico, M.; Facco, E.; Laio, A.; Rodriguez, A. Automatic topography of high-dimensional data sets by non-parametric density peak clustering. *Inf Sci.* **2021**, *560*, 476−492.

(26) Glielmo, A.; Macocco, I.; Doimo, D.; Carli, M.; Zeni, C.; Wild, R.; d'Errico, M.; Rodriguez, A.; Laio, A. DADApy: Distance-based analysis of data-manifolds in Python. *Patterns* **2022**, *3*, 100589.

(27) Rodriguez, A.; d'Errico, M.; Facco, E.; Laio, A. Computing the Free Energy without Collective Variables. *J. Chem. Theory Comput.* **2018**, *14*, 1206−1215.

(28) Facco, E.; d'Errico, M.; Rodriguez, A.; Laio, A. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. *Sci. Rep.* **2017**, *7*, 12140.

(29) Neyman, J.; Pearson, E. S. IX. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond.* **1933**, *231*, 289−337.

(30) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem* **2009**, *30*, 1615−1621.

(31) Gimondi, I.; Tribello, G. A.; Salvalaglio, M. Building maps in collective variable space. *J. Chem. Phys.* **2018**, *149*, 104104.
(32) Marinova, V.; Salvalaglio, M. Time-independent free energies from metadynamics via mean force integration. *J. Chem. Phys.* **2019**, *151*, 164115.
(33) 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.* **2004**, *25*, 1157−1174.
(34) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(35) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(36) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(37) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, *7*, 95−99.
(38) Hunnisett, L. M.; et al. The seventh blind test of crystal structure prediction: Structure ranking methods. *Acta Crystallogr., Sect. B* **2024**, *80*, 517−547.
(39) Sobornova, V. V.; Belov, K.; Krest'yaninov, M.; Khodov, I. Influence of Solvent Polarity on the Conformer Ratio of Bicalutamide in Saturated Solutions: Insights from NOESY NMR Analysis and Quantum-Chemical Calculations. *Int. J. Mol. Sci.* **2024**, *25*, 8254.
(40) Vega, D. R.; Polla, G.; Martinez, A.; Mendioroz, E.; Reinoso, M. Conformational polymorphism in bicalutamide. *Int. J. Pharm.* **2007**, *328*, 112−118.
(41) Coutsias, E. A.; Seok, C.; Dill, K. A. Using quaternions to calculate RMSD. *J. Comput. Chem.* **2004**, *25*, 1849−1857.
(42) Landrum, G. *RDKit, Open Source Cheminformatics*. RDKit, 2022.

https://doi.org/10.1021/acs.jctc.5c01247
J. Chem. Theory Comput. 2025, 21, 12466−12480