JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Understanding Ring Puckering in Small Molecules and Cyclic Peptides

Lucian Chan, Geoffrey R. Hutchison, and Garrett M. Morris*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The geometry of a molecule plays a significant role in determining its physical and chemical properties. Despite its importance, there are relatively few studies on ring puckering and conformations, often focused on small cycloalkanes, 5- and 6-membered carbohydrate rings, and specific macrocycle families. We lack a general understanding of the puckering preferences of medium-sized rings and macrocycles. To address this, we provide an extensive conformational analysis of a diverse set of rings. We used Cremer−Pople puckering coordinates to study the trends of the ring conformation across a set of 140 000 diverse small molecules, including small rings, macrocycles, and cyclic peptides.

By standardizing using key atoms, we show that the ring conformations can be classified into relatively few conformational clusters, based on their canonical forms. The number of such canonical clusters increases slowly with ring size. Ring puckering motions, especially pseudo-rotations, are generally restricted and differ between clusters. More importantly, we propose models to map puckering preferences to torsion space, which allows us to understand the inter-related changes in torsion angles during pseudo-rotation and other puckering motions. Beyond ring puckers, our models also explain the change in substituent orientation upon puckering. We also present a novel knowledge-based sampling method using the puckering preferences and coupled substituent motion to generate ring conformations efficiently. In summary, this work provides an improved understanding of general ring puckering preferences, which will in turn accelerate the identification of low-energy ring conformations for applications from polymeric materials to drug binding.
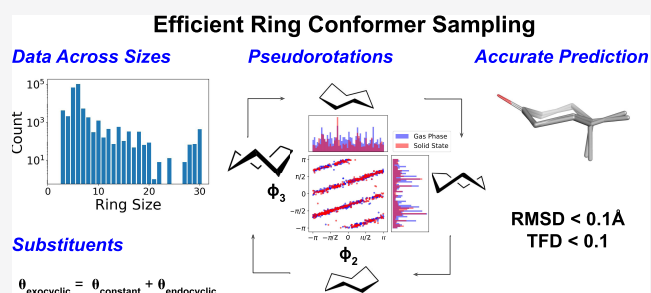
INTRODUCTION

Molecular rings play an important role in chemistry and biology, and their shapes are intimately linked to their physical and chemical properties. For instance, the glycosidase reactions heavily depend on their conformations.[1] Beyond small rings, macrocycle conformations are crucial in host−guest chemistry and drug design. In host−guest chemistry, the conformational preferences of macrocyclic rings lead to selective complexation of organic ligands.[2−4] On the other hand, macrocycles including cyclic peptides (CPs) have recently demonstrated their potential in modulating traditionally less druggable targets, e.g., mimicking protein−protein interactions.[5−8] The flexibility of cyclic molecules improves their chance to adopt favorable conformations that will bind to targets with flat surfaces. Despite the importance of ring conformations, most studies on ring conformations focus on small subsets, for example, on carbohydrate rings,[9−12] cycloalkanes,[13−15] and families of macrocycles involved in host−guest chemistry,[16−18] resulting in a lack of general understanding of ring conformational preferences, especially for medium-sized rings and macrocycles. We have therefore carried out an extensive conformational analysis on a wide range of ring molecules, including cyclic peptides.

Flexible rings can adopt different conformations due to out-of-plane bending motions, caused by changes in the rotatable ring bonds, resulting in so-called ring puckering. Typically, the ring puckers can be classified into different canonical forms and are usually low-energy conformations; a classic example is the chair and boat conformations in 6-membered rings such as cyclohexane. These canonical forms are not "unique", as the pseudo-rotation leads to multiple equivalent conformations, for example, the $^4C_1$ and $^1C_4$ chair conformations in cyclohexane.[10,19] The pseudo-rotation and the coupled change in substituent orientation sometimes lead to diverse geometry, i.e., large root-mean-square deviation (RMSD) in overall three-dimensional (3D) conformations, as illustrated in Figure 1. It is therefore necessary to sample ring conformations adequately to generate physically and biologically relevant conformational ensembles. In addition, there are several factors controlling the
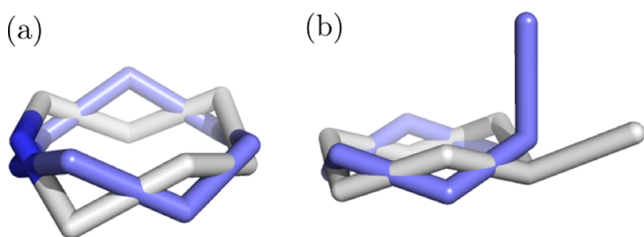
Figure 1. Two distinct pseudo-rotated conformations (white, lilac) of (a) azepane and (b) methylcyclohexane. The best RMSD between conformations are 0.60 Å and 0.67 Å respectively. The RDKit[23] implementation of the RMSD calculation was used. Pseudo-rotation and the concomitant change in substituent orientation, e.g., axial and equatorial methyl groups in panel (b), can lead to diverse geometries.

conformational flexibility of rings, including endocyclic double bonds,[20] the nature of substituents, and the presence of any intramolecular interactions such as hydrogen bonds.[21] Lyu et al. recently showed that the intramolecular hydrogen bonds restrict the pseudo-rotation path in deoxyribonucleosides, and the path characteristics depend on the strength of intramolecular interactions. In macrocycles, small structural modification, e.g., changes in exocyclic functionality, may lead to significant changes in conformation through emergent hydrogen bond and other intramolecular interactions.[22] Such conformational changes are difficult to predict, as the coupled ring bond rotations are not well understood.

A variety of coordinate systems have been developed to characterize ring puckers quantitatively. These techniques can be categorized into three general approaches. The first approach measures the perpendicular displacement of the ring atoms from a mean plane of the ring,[19,24] while the second approach makes use of the triangular tessellation of the ring and measures the associated angles between the reference plane and the triangular planes.[25] The last approach simply measures the ring torsion angles,[26] but this representation does not lend itself well to identifying pseudo-rotation. Methods used to analyze ring conformations based on perpendicular displacements of ring atoms such as Cremer–Pople puckering coordinates[19] are widely used in the community.[12,27] This representation has the advantage of using a reduced number of parameters, $N - 3$, to describe the geometry of an N-membered monocyclic ring. Hence, we require only two parameters to describe the conformational space of 5-membered rings and just three for 6-membered rings. It has also been used as collective variables for the enhanced sampling of 6-membered ring conformations in molecular dynamics studies.[28]

To better understand ring conformational preferences, we have extended our analysis to more complex ring systems, including larger sizes and bicyclic and polycyclic rings. We not only study their puckering preferences using Cremer–Pople puckering coordinates but also identify the underlying constraints on their geometry and the change in substituent orientations upon puckering. More importantly, we build quantitative models to convert from Cremer–Pople puckering coordinates to ring torsion angles, which thus allows us to understand the torsional changes upon pseudo-rotation. A novel knowledge-based conformational sampling scheme based on puckering parameters is also proposed. Unlike knowledge-based sampling methods, e.g., OMEGA,[29] which rely on a set of discrete prespecified ring templates and heuristic rules for sampling, our method efficiently explores conformational

space, including the dominant canonical conformations and their associated pseudo-rotation. We show that our sampling method can generate low-energy ring conformations effectively.

METHODS AND DATA

Cremer–Pople Puckering Parameters for *N*-Membered Rings. The out-of-plane deviations of puckered N-membered rings can be measured by the z-coordinates of the ring atoms relative to a mean plane cutting through the ring. The z-coordinates contain information about the overall movement or the shape of the puckered ring. Translation and overall rotation of the planar reference around the x- and y-axes can be removed by imposing three constraints (see Appendix 1, eqs S1–S3).

Let $\mathbf{R_j}$ be the position vectors of the ring atom, j, with the origin defined as the geometrical center of the puckered rings. We denote two vectors, $\mathbf{R}'$ and $\mathbf{R}''$, that define the mean plane (see Appendix 1, eqs S4 and S5), where $\mathbf{n}$ is the unit normal vector to this mean plane; z_j is then the displacement of atom j from the mean plane and is given by the scalar products in eq 1

$$z_j = \mathbf{R_j} \cdot \mathbf{n} \tag{1}$$

Using the mean plane and the full set of displacements, we can compute the Cremer–Pople ring puckering parameters[19] as follows.

For odd values of N, and $N > 3$, the puckering amplitude, q_m, and phase angle, ϕ_m, are defined as follows:

$$q_m \cos \phi_m = \left| \sqrt{\frac{2}{N}} \right| \sum_{j=1}^{N} z_j \cos\left(\frac{2\pi m(j-1)}{N}\right) \tag{2}$$

$$q_m \sin \phi_m = -\left| \sqrt{\frac{2}{N}} \right| \sum_{j=1}^{N} z_j \sin\left(\frac{2\pi m(j-1)}{N}\right) \tag{3}$$

eqs 2 and 3 apply for $m = 2, 3, ..., (N - 1)/2$. The amplitudes, q_m, are positive-valued, while the phase angles, ϕ_m, range from $-\pi$ to π radians.

For even values of N, eqs 2 and 3 apply for $m = 2, 3, ..., (N/2 - 1)$, but an additional puckering amplitude is required, with the following form:

$$q_{N/2} = \left| \sqrt{\frac{1}{N}} \right| \sum_{j=1}^{N} (-1)^{j-1} z_j \tag{4}$$

Note that the $q_{(N)/(2)}$ value in eq 4 can take either sign.

The Cremer–Pople representation is only applicable to monocyclic ring systems. To extend it to more complex ring systems such as fused and spiro rings, we first decompose the ring systems into smaller rings and calculate the puckering parameters for each ring. In particular, we adopt the concept of unique ring families (URFs)[30] for this decomposition, with the resultant Cremer–Pople parameters being calculated for all relevant cycles, i.e., minimum cycle bases.

Additionally, the Cremer–Pople representation is atom-order-dependent, and we standardize the atom ordering in the ring before calculation. Bond orders, connectivity, and element types are used to determine this standardized order (see Appendix 1). Other canonical atom numberings may also be used.[31] In symmetric rings, such as cycloalkanes, the first atom is picked at random.

We also take the volume of the amino acid into account when ordering the backbone ring atoms in cyclic peptides. The priority increases with volume, so tryptophan, tyrosine, and phenylalanine have higher ranks, while glycine has the lowest. The rank order of amino acids can be found in Appendix 1, Table S1. Note that this ordering is only applied to the cyclic peptides.

Extension of Cremer–Pople Puckering Parameters for Ring Substituent Positions. It is well known that the preferred orientation of ring substituents changes under ring inversion, and neighboring substituents can also influence their preferred orientation. We followed the framework proposed by Cremer[32] to describe the position of substituents unambiguously. Two orientation angles, α and β, are introduced. The α angle describes the relative position of the substituent to the mean plane defined above (see Appendix 1, eq S11), and β angle describes the relative position of the substituent to the geometrical center of the ring (see Appendix 1, eqs 12 and 13; see Figure 2 for illustration).
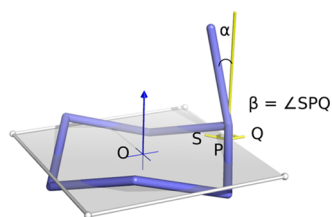
Figure 2. Definition of the substituent orientation angles α and β. Methylcyclohexane is used as an example, with a mean plane (gray) cutting through the 6-membered ring. The methyl substituent is axial to the mean plane ($\alpha = 0.24$ rad). O denotes the origin, which is also the geometrical center of the ring. The points S and P are projections of the methyl carbon and the ring atom that is attached to the methyl carbon onto the mean plane. The point Q lies in the mean plane such that points O, P, and Q are collinear. The angle β is defined by the angle between S, P, and Q, and $\beta = -2.25$ rad in this example.

When the substituent angle $\alpha = 0$ or π, this indicates that the substituent is sitting axially above or below the mean plane, respectively, while $\alpha = \pi/2$ indicates the equatorial orientation. The angle $\beta = 0$ indicates a radially outwardly directed substituent, while $\beta = -\pi$ or $\beta = \pi$ indicates an inwardly directed substituent.

With this complete representation for the ring puckering motion and substituent orientation, we can investigate their coupled motion extensively and develop ring puckering potentials for conformer sampling, similar to their acyclic counterparts.

Connection between Ring Puckering, Substituent Orientations, and Torsion Angles. Ring inversion is the interconversion of cyclic conformers that have equivalent ring shapes. Such interconversion can be characterized by Cremer–Pople representation. The substituent orientation also changes during inversion. In particular, we are interested in the coupled ring bond rotations and the associated change in substituent orientation during pseudo-rotation. Inspired by the functional forms studied in previous work,[33] three models are proposed (see Appendix 1, eqs S18–S20). Equation S18 is used to predict the associated change in substituent α and β orientation angles upon puckering. Equation S19 maps the puckering parameters to endocyclic torsion angles, while eq S20 helps explain the rotational dependence between the substituent exocyclic torsion angle and endocyclic torsion angle. Note that eq S19 is a mapping for the general N-membered ring, and the functional form proposed by de Leeuw et al.[33] to convert puckering coordinates to torsion angles for 5-membered rings can be recovered by applying trigonometric identities.

Here, we denote the endocyclic torsion angle as θ^{endo}; the exocyclic torsion angle as θ^{exo}; and α and β as the substituent orientation angles.

Ring Reconstruction from Cremer–Pople Puckering Parameters. Cremer–Pople puckering parameters not only provide quantitative descriptions of puckered N-membered rings but also allow efficient conversion from puckering parameters to Cartesian coordinates, as shown by Cremer.[34] In addition to $N - 3$ puckering parameters, $N - 3$ bond angles and N bond lengths are required for the reconstruction of puckered N-membered ring conformations. The default values of bond lengths and bond angles are specified in Tables S2 and S3 in Appendix 1. The calculation of the x-, y-, and z-coordinates from puckering parameters, specified bond lengths, and bond angles is discussed in Appendix 1.

To sample low-energy ring conformations efficiently, we used kernel density estimation (KDE) to learn the ring puckering preferences and generate puckering values from the model. Note that the Cremer–Pople parameters were mapped to Cartesian space ($q_m \cos \phi_m$, $q_m \sin \phi_m$) for the KDE calculation. A Gaussian kernel was used for the density estimation. The samples were then converted to different z-coordinates to give distinct ring conformations. Using the relationship between endocyclic torsion angles and exocyclic torsion angles (see Appendix 1, eq S20) with appropriate parameters (see Appendix 2, Table S8), we can update the ring substituent position accordingly. Note that the exocyclic bond angles are kept fixed in the sampling. This approach is in contrast to traditional knowledge-based sampling methods,[29,35,36] where ring templates and heuristic rules are used to sample ring conformations, and substituent positions are then assigned by minimizing a clash function or force field energy. Our approach does not require force field minimization, although as discussed below, minimization can also improve cases where the actual bond lengths or angles differ slightly from our model.

Data. Over 130 000 small molecules were selected from the Crystallography Open Database (COD)[37,38] (63814 molecules) and the ZINC database[39] (67009 molecules), including natural products and macrocycles. Molecules from COD and ZINC contain hydrogen, boron, carbon, nitrogen, oxygen, fluorine, silicon, phosphorus, sulfur, chlorine, bromine, and iodine. Molecules with carbon, nitrogen, oxygen, and sulfur in a ring with up to 20 atoms were considered. For COD molecules, the Open Babel version 2.4[40] was used to convert from CIF format to SDF format and assigned bond orders. Molecules with inconsistent geometries, such as hydrogen atoms or consecutive double bonds contained in a ring, were excluded from the analysis. In addition, we generated a set of cyclic peptides (CPs), including 8661 cyclic tetrapeptides (CTPs) and 2249 cyclic pentapeptides (CPPs). The peptide data sets contain head-to-tail cyclic tetrapeptides and cyclic pentapeptides, i.e., cyclization from the N-terminus to the C-terminus, yielding a set of 12-membered and 15-membered rings. Their sequences are composed of 14 of the 20 naturally occurring L-amino acids (see Appendix 3, Table S9).

For all molecules from ZINC and the cyclic peptides, experimental-torsion distance geometry with basic knowledge[41] was used to generate initial geometry, followed by geometry optimization using the GFN2 method[42] and conformer sampling using the iterative metadynamics sampling and genetic crossover (iMTD-GC) method implemented in the CREST program.[43,44] Note that this data set is also used in our previous works.[45] We should note that CREST may break the molecules into smaller fragments in the output file. Such fragmented molecules were excluded from our analysis.

To demonstrate the effectiveness of using puckering preferences in sampling ring conformation, we selected 20 simple molecules, including monocyclic rings with and without endocyclic double bonds and substituents (see Appendix 3, Table S10).

Analysis. To provide a better understanding of the ring geometry in cyclic peptides, we computed the (ϕ, ψ) torsion angles. We also calculated the eccentricity, which is used to measure the "roundness" of a ring.[46] Eccentricity, e, is a nonnegative real value that characterizes the shape of a conic section. A value of 0 indicates a circle and 1 indicates an ellipse.

To assess the performance of our proposed sampling method, we computed the heavy atom root-mean-square deviation (RMSD) and torsion fingerprint deviation (TFD)[47] between the generated conformations and the lowest-energy (reference) conformation sampled from CREST.

Furthermore, three metrics, namely, squared circular correlation coefficient (R^2_{circ}), mean angular error (MAE), and standard deviation of the angular error, were used to assess the predictive performance of our proposed models. The circular correlation coefficient and the angular error (circular distance between the predicted and actual angles) are defined by eqs S22 and S23 in Appendix 1, respectively.

Implementation. RDKit[23] was used to read molecules, generate conformations, and write conformers. The implementations of RMSD and TFD calculation in RDKit were used. RingDecomposerLib[48] was used to identify the URFs of the ring system. The implementation of KDE in Scikit-Learn[49] was used. The code is available in Github (https://github.com/lucianlschan/RING).

■ RESULTS AND DISCUSSION

Small- and Medium-Sized Rings. A relatively small number of conformational clusters were observed for 5- to 8-membered rings, reflecting their canonical conformations. For instance, Figure S4a in Appendix 3 shows two clusters for flexible 6-membered rings, corresponding to the celebrated chair and boat conformations, as illustrated in Figure 3. As expected, the chair conformation is more frequently observed than the boat conformation. The phase angle, ϕ_2, is uniformly distributed, suggesting free pseudo-rotation in both forms. In contrast, the presence of endocyclic double bonds or shared aromatic bonds restricts both puckering and pseudo-rotation. The puckering amplitude, q_3, and phase angle, ϕ_2, exhibit a sinusoidal relationship as can be seen in Figure S4b in Appendix 3. These relationships hold for both simple monocyclic rings and complex bi- and polycyclic rings.

For 7- and 8-membered rings, an additional phase angle, ϕ_3, is required. Phase—phase couplings are evident in some conformational clusters. For example, three conformational clusters were observed in 7-membered rings with no endocyclic double bonds, having predominantly twist-chair and chair conformations, as illustrated in Figure 4a. The
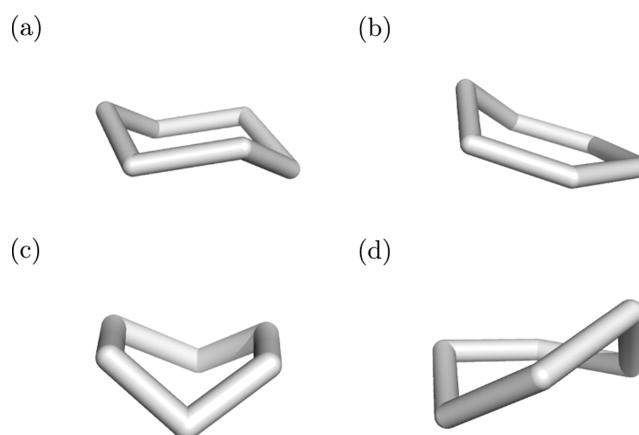
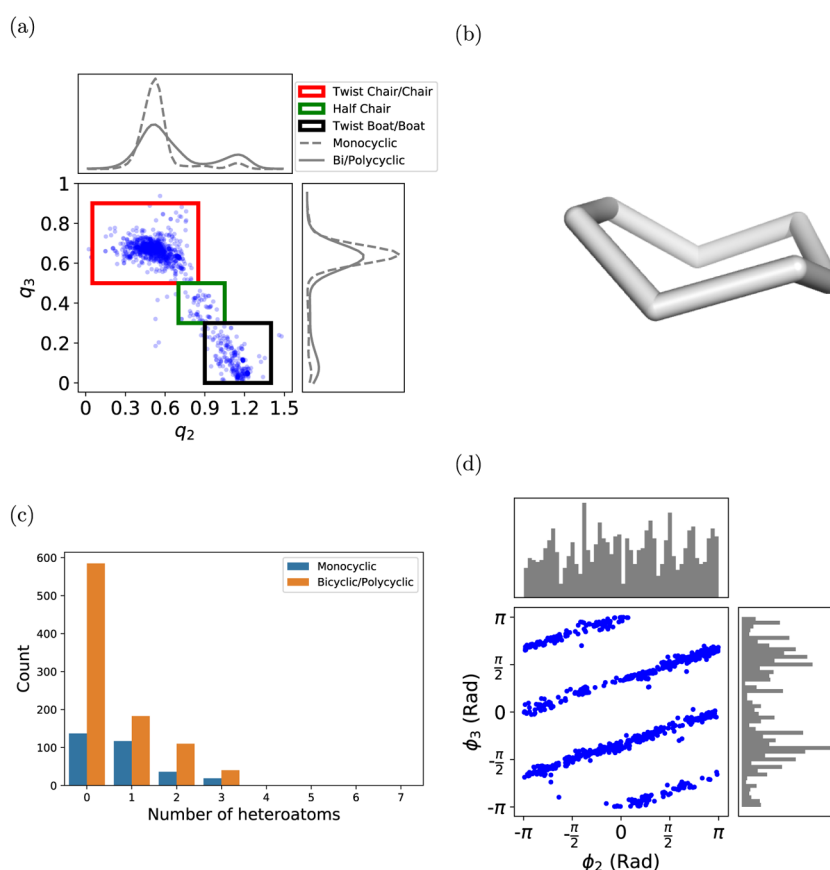**Figure 3.** 6-membered ring conformations: (a) chair, (b) half-chair, (c) boat, and (d) twist boat.

puckering amplitudes (q_2, q_3) fall into a narrow range, and the pseudo-rotations are restricted in this region, as shown in Figure 4c. The phase angles ϕ_2 and ϕ_3 are strongly coupled, and they are marginally uniformly distributed. This coupled motion suggests the minimum energy pathway of the chair—twist-chair pseudo-rotation. As suggested by Bocian et al.,[13] the pseudo-rotation map can be approximated by eq 5, with varying intercepts (ϕ_2^*, ϕ_3^*) and slopes (K_2, K_3). This model is valid for all rings with heteroatoms (see Figure 4b).

$$\phi_2 = \phi_2^* + K_2\phi$$

$$\phi_3 = \phi_3^* + K_3\phi \qquad (5)$$

In bicyclic and polycyclic rings, the adjacent rings and bulky substituents sometimes induce significant steric clashes and result in concomitant changes in conformational preferences. The increase in amplitude q_2 and decrease in amplitude q_3 indicate a conformational change from chair to half-chair ($0.7 < \phi_2 < 1$) and boat conformations ($\phi_2 > 1$). The pseudo-rotations are free in these clusters, i.e., the phase angles are randomly distributed (see Appendix 3, Figure S5c).

To assess the effect of the endocyclic double bonds on conformational preferences, we selected 7-membered rings with one and two endocyclic double bonds. We further separated the observations by the location of endocyclic double bonds. Figure S6 in Appendix 3 shows three conformational clusters in 7-membered rings with single endocyclic double bonds, and they correspond to the chair, half-chair, and boat conformations, which are the same as the case without double bonds. However, the population of the chair conformation decreases, while the population of half-chair and boat conformations increases. The pseudo-rotations in all three clusters are restricted. In the chair and twist-chair regions, the phase angle, ϕ_3, is relatively fixed with small variations in the phase angle, ϕ_2, while in the boat and twist boat regions, the phase angle ϕ_2 is fixed while the phase angle ϕ_3 varies. The half-chair conformation exhibits strong coupling between phase angles.

As the number of endocyclic double bonds increases, the number of degrees of freedom of the ring system decreases. The location of the double bonds strongly influences the puckering preferences, as shown in Figure 5. The double bonds in 1,3-cycloheptadiene and 1,4-cycloheptadiene-like structures (Figure 5a,5c) impose different steric constraints and lead to

Figure 4. Analysis of 7-membered rings with no endocyclic double bonds. (a) Joint distribution and marginal distribution of the ring puckering amplitudes (q_2, q_3). This shows that twist-chair and chair conformations (indicated by a red box) are frequently observed in the lowest-energy conformation, followed by boat and twist boat conformations (indicated by a black box). The half-chair (indicated by a dark green box) is the transition structure from chair to boat, and it is occasionally observed. The shape of monocyclic rings is conserved, while there is some variation in bicyclic and polycyclic rings. Note that the color boxes only show the coarse boundary of the conformational clusters. (b) Example of the chair conformation found in cycloheptane (hydrogen atoms not shown). (c) Histogram showing the count of molecules with varying numbers of heteroatoms in rings found in the chair or twist-chair conformation. (d) Coupled phase angles of the chair and twist-chair conformations, as indicated by the red box in (a). This plot reveals the minimum energy pseudo-rotation pathway of the chair and twist-chair conformations. This relationship holds for general 7-membered rings with or without heteroatoms.

contrasting phase−phase coupling. The correlations in amplitudes are shown in Appendix 3, Figure S7.

For larger rings, the number of conformational clusters increases, while the coupling between puckering amplitudes and phase angles becomes more complex. It should be noticed that small local structural changes may result in significant changes in conformation through transannular repulsion and intramolecular interactions. To gain further insight into long-range-coupled ring bond rotations, we performed cluster analysis on a set of cyclic peptides.

Cyclic Peptides. Peptide cyclization imposes additional constraints on the system and thus reduces the thermally accessible conformational space of the resultant cyclic peptides relative to their linear counterparts.[50] There are several factors governing the backbone conformation of cyclic peptides, including the size and properties of the amino acid side chains, the presence of N-methylation, and the formation of γ- and β-turns. Analyzing the puckering preferences helps understand the relative influence of these factors.

The configuration of the amide bonds provides important information to determine the dominant backbone conformation adopted by the cyclic peptides. The partial double bond character of the carbon−nitrogen bond in amide bonds renders

them planar, resulting in either *cis* (C) or *trans* (T) amides. We can thus classify the conformations based on the sequence of *cis*- or *trans*-amide bonds, as described in Loiseau et al.,[51] for example, for cyclic tetrapeptides, all-*cis* ("CCCC") or all-*trans* ("TTTT") amides. Typically, the *trans*-amide bond is preferred in acyclic peptides, large cyclic peptides, and proteins. Figures S9a and S13a in Appendix 3, however, show that the *cis*-amide bond is preferred in both cyclic tetrapeptides and cyclic pentapeptides, with 40% all-*cis* and 43% CCCT in cyclic tetrapeptides. In small cyclic peptides, high ring strain reduces the energy barrier between *cis* and *trans* isomers. All-*trans* and single-*cis* (CTTT and CTTTT) configurations are less favored in both tetra- and pentapeptides due to high transannular strain, and they exist only with explicit stabilization from one or more intramolecular hydrogen bonds. Such stabilization leads to γ-turns in cyclic tetrapeptides and γ- and β-turns in cyclic pentapeptides, as reflected by their Ramachandran (ϕ, ψ) dihedral angles: see, for example, Appendix 3, Figure S12a. The puckering amplitudes and phase angles are thus highly restricted in such conformational clusters. It should be noted that these turns are favored by the *in vacuo* calculation and may not reflect the conformations observed in solution. The positional preferences of amide carbonyl groups are key to
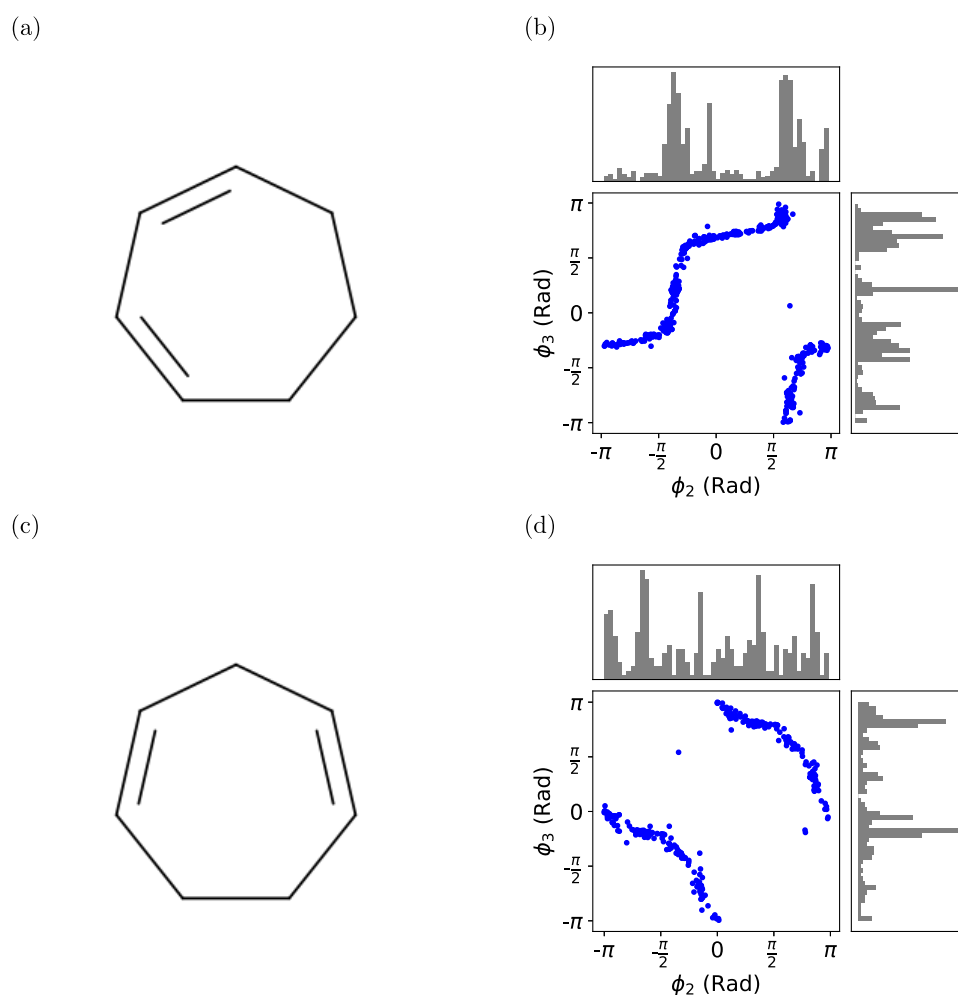
Figure 5. 7-membered rings with two endocyclic double bonds and their associated phase angle coupling. (a) 1,3-Cycloheptadiene, and (b) the highly coupled phase angles of the low-energy conformations observed in 7-membered rings with double bonds at the 1 and 3 positions. (c) 1,4-Cycloheptadiene and (d) again, the highly coupled phase angles of the low-energy conformations observed in 7-membered rings with double bonds at the 1 and 4 positions. Monocyclic, bicyclic, and polycyclic rings are all included in our analysis, and it should be noted that the double bond can also be a shared aromatic bond. The relative location of the endocyclic double bonds imposes different constraints on the system and results in visibly different phase−phase couplings.

understanding the formation of such intramolecular hydrogen bonds, which we discuss next.

Main chain−main chain intramolecular interactions were not observed in cyclic tetrapeptides with two or more *cis*-amide bonds, nor were they seen in cyclic pentapeptides with three or more *cis*-amide bonds. Transannular repulsion, main chain−side-chain, and side-chain−side-chain intramolecular interactions appear to be the major driving forces behind the conformational preferences seen in these cases. Small structural modifications, such as the change in amide bonds and/or side-chain orientations, may induce significant steric clashes and lead to conformational switching. For example, Figure 6a,6b shows the puckering amplitude preferences of two canonical conformations in all-*cis*-amide cyclic tetrapeptides, and they differ by the orientation of one amide bond. Similarly, we followed the nomenclature used in Loiseau et al., where the orientation of amide carbonyl is denoted by *U* when it is oriented above the mean plane, while it is denoted by *D* when it is oriented below the mean plane. The two canonical forms (CCCC−DDDD and CCCC−UDDD) exhibit distinct puckering amplitude preferences and phase−phase couplings (see Appendix 3, Figure S10). Similar phenomena are observed in

cyclic pentapeptides (see Appendix 3, Figure S14). Furthermore, the formation of main chain−side-chain interactions and/or side-chain−side-chain interactions give rise to two subclusters within the same configuration (CCCC−DDDD) with diverse geometries, as illustrated in Figure 7a,7b. The orientation of the side-chain C_β atoms plays important roles in the formation of these interactions.

To further understand the cyclic backbone conformation, we calculated the Ramachandran (ϕ, ψ) dihedral angles and the eccentricity of the backbone. The Ramachandran plots in Appendix 3, Figures S12b, and S13d show that the (ϕ, ψ) angle preferences of cyclic tetrapeptides and pentapeptides are similar to those of the standard secondary structures observed in proteins. Figures S9b and S13b show contrasting eccentricity values between clusters; for example, all-*trans* cyclic tetrapeptides give a mode at 0.3, while alternating CTCT cyclic tetrapeptides give a mode at 0.8, indicating diverse geometries between clusters.

We thus have shown that Cremer−Pople puckering parameters are a useful representation to understand ring puckering for both small rings and macrocycles including cyclic peptides and analyzed the associated effects of endocyclic

(a)

(b)
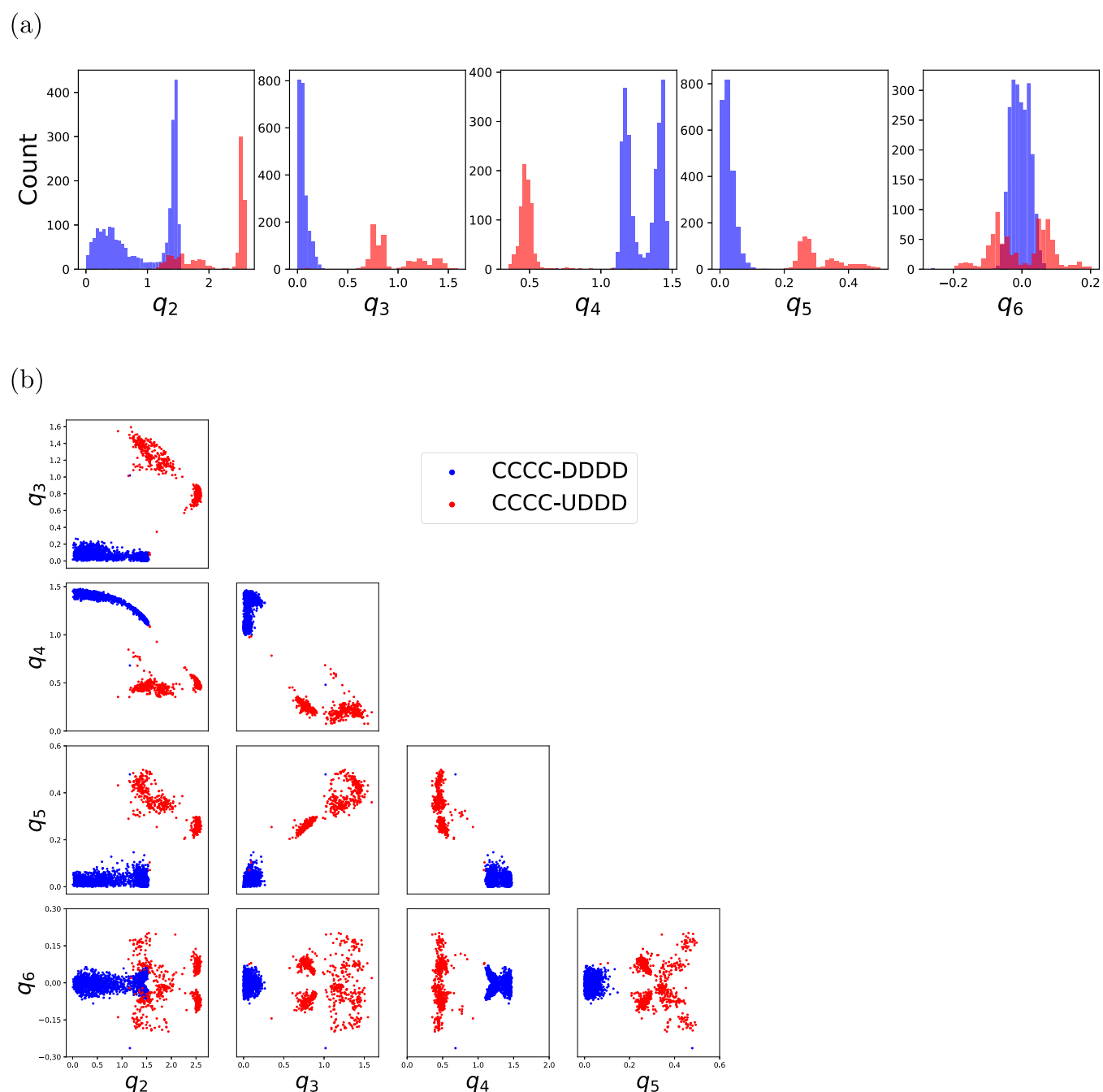
Figure 6. (a) Marginal distribution of the ring puckering amplitude (q_2, q_3, q_4, q_5, q_6) preferences for two conformational clusters of all-*cis* conformation in cyclic tetrapeptides (colored red and blue). The two clusters are defined by the α orientation angle of the amide carbonyl oxygen, where U indicates $\alpha < \pi/2$, and D indicates $\alpha > \pi/2$. The CCCC−DDDD conformations are colored blue, while CCCC−UDDD are colored red. In panel (a), two modes are observed in puckering amplitudes for both clusters, indicating the presence of multiple subclusters. (b) Pairwise joint distribution of the ring puckering amplitude (q_2, q_3, q_4, q_5, q_6) preferences for two conformational clusters of all-*cis* conformation in cyclic tetrapeptides. The puckering preferences of CCCC−DDDD conformations are more concentrated than those in CCCC−UDDD conformations.

double bonds on ring puckering. We have also revealed the influence of configuration and orientation of amides on ring geometries. To gain further insights, we will examine the substituent orientations and their relationship to puckering preferences below.

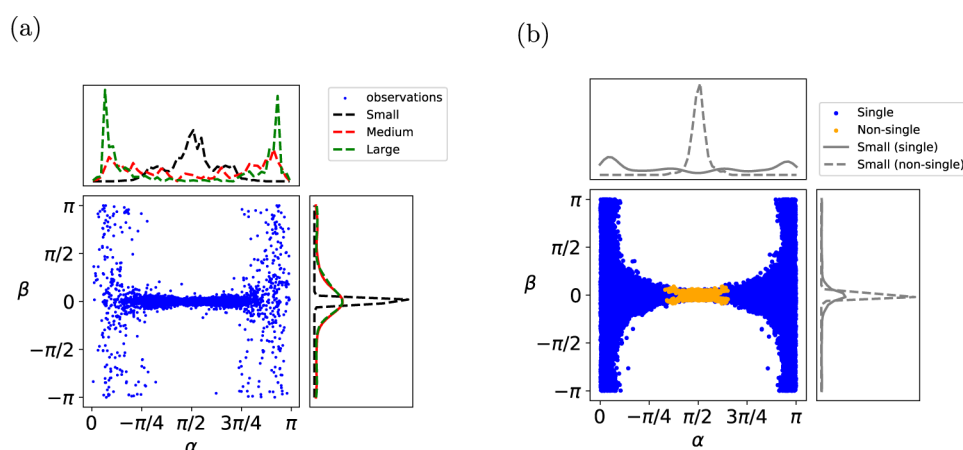**Effects of Substituent Orientation and Functionality.** The size and functionality of substituents are two of the key factors determining the ring geometries, and their effects vary with ring size. We thus separated the lowest-energy conformations according to ring sizes: small (5- and 6-

membered) rings, medium (7- to 11-membered) rings, and macrocycles (12-membered or larger rings).

As might be expected, ring substituents tend to be outwardly directed (relative to the ring center) in small- and medium-sized rings, i.e., β is close to zero, regardless of the nature of the substituents (see Appendix 3, Figure S15). However, substituents including carbonyl and hydroxyl are allowed to be quasi-axial to the mean plane and inwardly directed in macrocycles, which are sterically unfavorable in small- and medium-sized rings. Their α angle preferences, however,
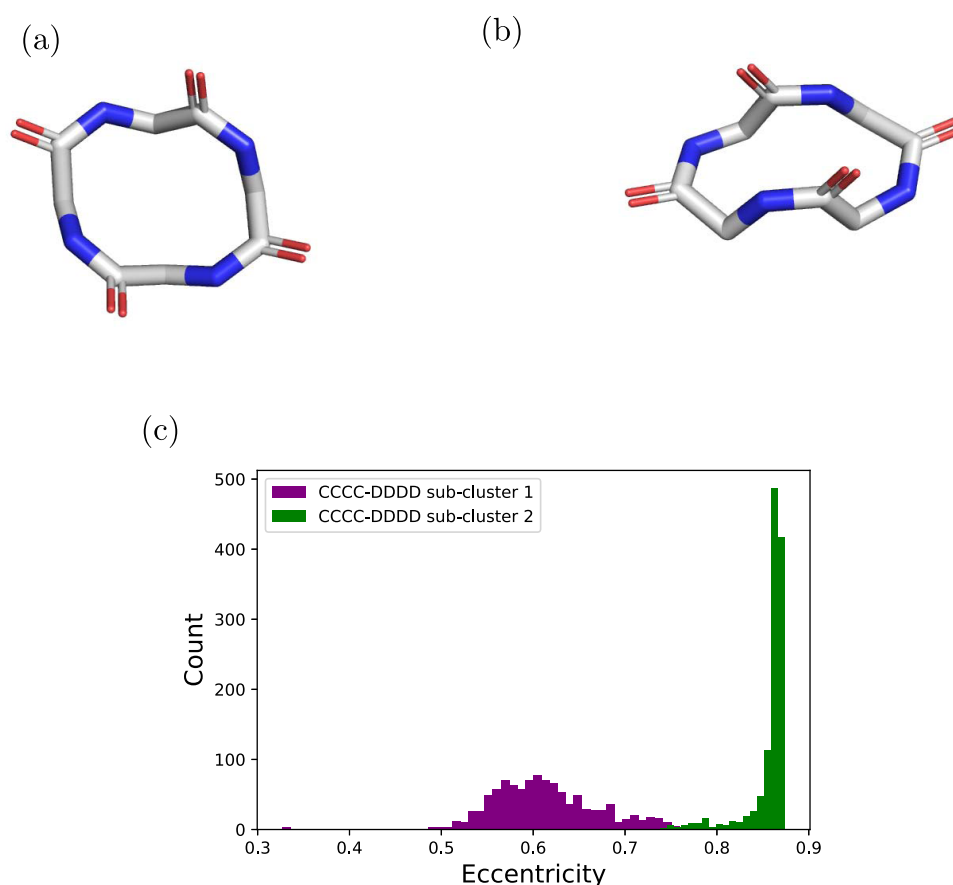
Figure 7. Example conformation from (a) subcluster 1 and (b) subcluster 2. Hydrogen atoms and side chains are not shown in panels (a) and (b). (c) Ring eccentricity values for two subclusters of CCCC−DDDD conformations are colored purple (subcluster 1) and green (subcluster 2). The main chain−side-chain and side-chain−side-chain intramolecular interactions give rise to diverse geometries.
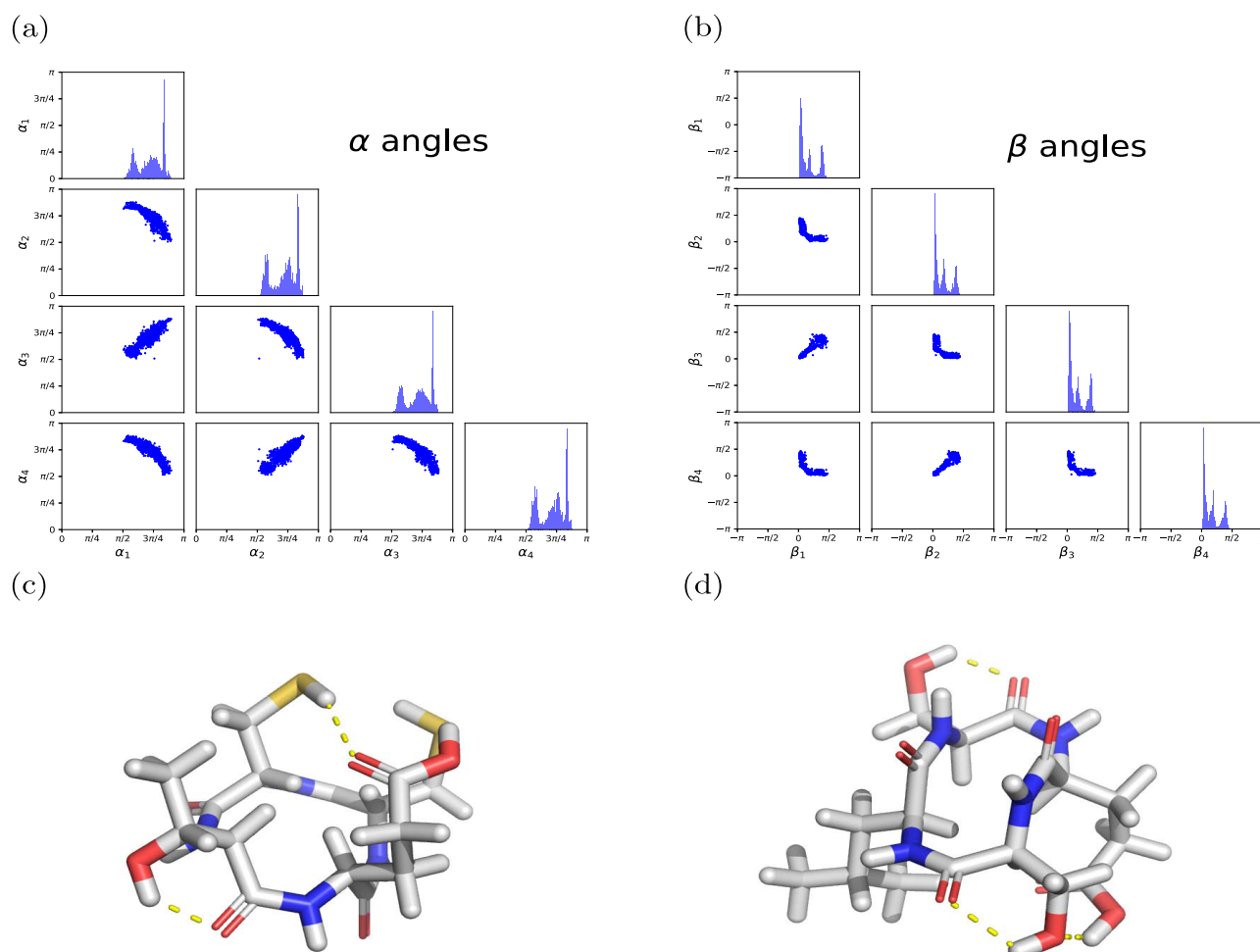

Figure 8. Substituent orientation angle preferences for (a) carbonyl functional groups and (b) methyl functional groups, attached to small rings with and without endocyclic double bonds. Panel (a) shows that the carbonyl groups tend to be equatorial to the mean plane ($\alpha \approx \pi/2$) in small rings, while panel (b) shows that the orientation of a methyl group tends to be restricted when it is attached to a ring atom that is linked to a neighboring ring atom with a shared endocyclic double bond.

depend on both ring size and the nature of substituents. For example, Figure 8a shows the substituent orientation preferences of the carbonyl functional group. Due to the exocyclic double bond, its movement is restricted compared to other single-bonded small substituents such as hydroxyl and methyl. The carbonyl oxygen thus tends to be equatorial to the

mean plane, i.e., $\alpha \approx \pi/2$ in small rings, and preferences change as the ring size increases. Besides exocyclic double bonds, endocyclic double bonds also restrict the exocyclic motion. Figure 8b shows the substituent orientation preferences of methyl groups in small rings, and the α angle is bounded when the methyl is attached to a ring atom that is linked to a

Figure 9. Substituent orientation angle preferences of amide carbonyl groups in cyclic tetrapeptides with CCCC−DDDD conformations, where C denotes *cis*-amides and D indicates $\alpha > \pi/2$, where α is the orientation angle of the ring "substituent" amide carbonyl oxygen: (a) preferred α angles and (b) preferred β angles. Both sets of plots show strong coupling motion between each of the four backbone carbonyl oxygen atoms to avoid steric clashes and/or align main chain−side-chain and side-chain−side-chain intramolecular interactions. An example with the main chain−side-chain interaction and side-chain−side-chain interactions (yellow dotted line) in (c) subcluster 1 and (d) subcluster 2.

neighboring ring atom with a shared endocyclic double bond. The influence of endocyclic bonds is weakened in medium-sized rings and macrocycles, and the α angle can therefore adopt a wider range of values.

To reveal the role of bulky substituents in macrocycles, we assessed their orientation angles in cyclic peptides, in particular, the positional preferences of amide carbonyls and the side-chain C_β atoms. As mentioned above, there are multiple conformational clusters in cyclic peptides. In particular, γ-turns are observed in the all-*trans* and CTTT conformation in cyclic tetrapeptides, and the formation of main chain intramolecular hydrogen bond leads to a rigidification of amide carbonyl positions, as illustrated in Appendix 3, Figure S17. On the other hand, amide carbonyl groups in other clusters move accordingly to avoid steric clashes and/or align main chain−side-chain intramolecular interactions (see, for example, Figure 9). Likewise, the C_β atoms of all amino acids studied except Gly show correlated motions, so as to avoid steric clashes and align side-chain−side-chain interactions. In addition to C_β orientation, we calculated the side-chain torsion angles, χ_1. Figure S18 in Appendix 3 shows multimodality in χ_1 angles, which is consistent with side-chain torsion angles observed in protein

secondary structures. This suggests that the side-chain conformations can be easily sampled using standard side-chain rotamer libraries.[52]

The extended Cremer−Pople representation provides a means to understand correlated positional preferences in ring substituents; however, it is not clear what the relationship between the puckering preference and substituent orientation is, especially in macrocycles. We have therefore developed simple models (Appendix 1, eq S18) to predict α and β orientation angles. Figure 10a,b shows the predictions of the α and β orientation angles of carbonyl groups in a 6-membered ring at the given positions. The predicted values are in good agreement with the actual values, with low mean angular error and high squared circular correlation coefficient. The model is also valid for larger rings.

Relationship between Ring Puckering Parameters, Substituent Orientations, and Torsion Angles. As mentioned earlier, measuring torsion angles is an alternative way to quantify ring puckering and is often used in conformational analysis of small rings; de Leeuw et al.[33] discussed the connection between ring puckering coordinates and torsion angles for small rings. Here, we proposed a general model defined in Appendix 1, eq S19, to convert puckering
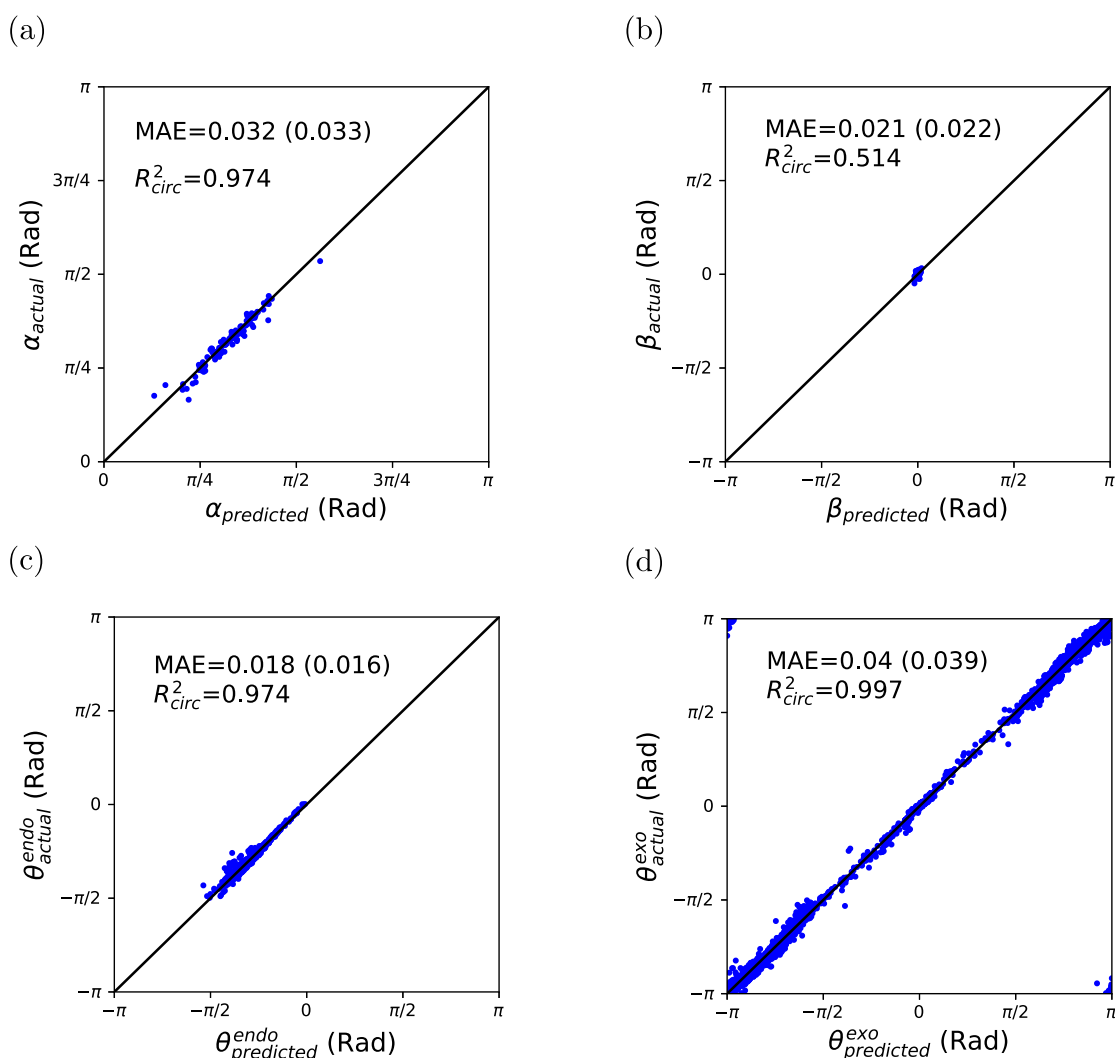
Figure 10. Predictions of the (a) α and (b) β orientation angles of a carbonyl group at position 1 in a 6-membered ring chair conformation. The mean angular errors (MAEs), standard deviation of angular errors (shown in parentheses), and the squared circular correlation coefficients are reported. The low squared circular correlation in panel (b) is ascribed to the rigidity of β orientation angles in small rings. (c) Predictions of the first endocyclic torsion angles in a 6-membered ring chair conformation. The predictive performance of other endocyclic torsion angles can be found in Appendix 2, Table S7. (d) Predictions of exocyclic torsion angles of a carbonyl group in all positions for ring sizes up to 16. All proposed models show excellent agreement with the actual substituent orientation angles and torsion angles, with low mean angular errors and high squared circular correlation coefficients.

parameters to endocyclic torsion angles for N-membered rings. Figure 10c shows the predictions of all endocyclic torsion angles of 6-membered rings. All position submodels show good agreement with the actual ring torsion angles, with high squared circular correlation coefficient values, $R^2_{\text{circ}} > 0.9$. This model is also valid for larger rings. The improved understanding of how the rings switch conformations and pseudo-rotate enables the use of metadynamics simulation with appropriate coordinates to effectively sample the conformational space of macrocycles and cyclic peptides.[53]

Equation S20 in Appendix 1 defines the relationship between the change in substituent exocyclic torsion angles $(s_i, i, i + 1, i + 2)$ with respect to the neighboring endocyclic torsion angles. $(i - 1, i, i + 1, i + 2)$, where s_i is the substituent atom (say, a carbonyl oxygen) attached to its ring atom i (say, a carbonyl carbon), and $i + k$ $(k = -1, 1, 2)$ are the ring atom positions. Figure 10d shows the excellent fit between the predicted and actual exocyclic torsion angles of carbonyl groups at different positions, regardless of the ring size. Our

model gives a high squared circular correlation coefficient, R^2_{circ} = 0.997, and a small mean angular error, 0.04 radian ($\approx 2.3°$). Similar performance can be achieved for other substituents. These models allow us to assign substituent positions efficiently once the ring conformation is defined. We should note that the exocyclic bond angles will also change upon puckering, but their relationship with ring puckering parameters is not discussed here.

Puckering and Substituent Orientation Preferences in Solid State. We have so far presented the gas-phase puckering preference, using *in vacuo* GFN2 energy evaluations. To gain insights into the puckering preference in solid state, we compared our results with previous empirical studies on crystal structures from the Cambridge Structural Database[54−56] and the 63814 experimentally determined X-ray crystal structures from COD. These empirical studies focused on medium-sized rings, in particular, 7- and 8-membered rings, and showed similar puckering preferences and pseudo-rotations of the dominant canonical conformations. However, the actual

distributions slightly differ from our work due to the small number of crystal structures used in their studies. Similarly, Figure S19 in Appendix 3 shows that the puckering preferences for small- and medium-sized rings are similar in both gas phase and solid state. The coupling between the substituent orientation angle and ring puckering is almost identical to that in GFN2-computed low-energy structures. Our proposed models can therefore be applied directly to the solid-state conformation. For larger rings, we are not yet able to make any conclusions due to limited numbers of observations. These results suggest that cyclic small molecules generally adopt low-strain conformations in both the solid state and gas phase.[57] Likewise, the intramolecular and intermolecular interactions can be aligned by pseudo-rotation or conformational switching (from one canonical form to another canonical form) in the solid state.

Ring Reconstruction. We selected 20 simple ring systems, including monocyclic rings with and without substituents and endocyclic double bonds, to assess the performance of our proposed sampling method based on the puckering preference. Note that these molecules do not contain any acyclic rotatable bonds. The lowest-energy conformations from CREST, calculated *in vacuo* using the GFN2 energy function, were used as reference conformations. The lowest RMSD value and the TFD value of the corresponding conformation were reported. Figure 11 shows two examples, cycloheptane and 4,4-
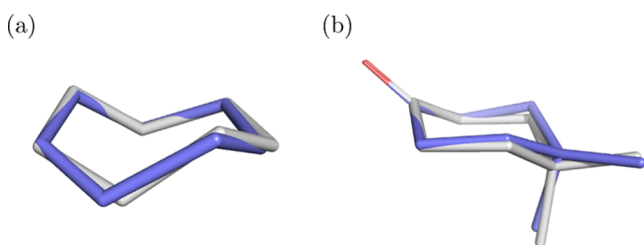
Figure 11. Alignment of conformations generated by our method (in lilac) and the lowest-energy conformation sampled by CREST (in white) for (a) cycloheptane and (b) 4,4-dimethylcyclohexanone. The sampled conformations are very similar to the lowest-energy conformation, with RMSD values of 0.12 Å and 0.16 Å and TFD values of 0.06 and 0.05, respectively. The torsion deviations are small in both cases, and the deviation in bond lengths and bond angles leads to larger RMSD values. RDKit[23] was used to compute the RMSD and TFD values. Figures were generated using PyMOL.[58]

dimethylhexanone. Both have generated conformations (without energy minimization) that are very similar to their corresponding reference conformations, with low RMSD values (0.12 Å and 0.16 Å, respectively) and TFD values (0.06 and 0.05, respectively). In general, our proposed method gives low average TFD values (0.05) and an average RMSD value of 0.09 Å on the selected cyclic molecules. This demonstrates the effectiveness of our proposed method. Note that the large RMSD values are ascribed to the deviation in bond lengths and bond angles. Local geometry optimization with bond lengths and bond angles will help generate a better conformation with lower RMSD values.

CONCLUSIONS

We have investigated the ring puckering motions of over 140 000 flexible cyclic small molecules and cyclic peptides (CPs) using Cremer−Pople puckering parameters. By standardizing the atom ordering of the ring atoms, we have been

able to elucidate the coupled motions and torsional preferences for *N*-membered ring molecules from GFN2-computed low-energy structures. The representation can be easily extended to describe substituent geometries unambiguously, thus enabling us to study the coupled motion of ring substituents upon puckering.

We have shown that the presence of endocyclic double bonds and shared bonds with aromatic rings constrains the ring system and results in a corresponding change in ring puckering. In addition, the pseudo-rotations are generally restricted. The pseudo-rotation is only "free" in some conformational clusters, e.g., the chair conformation in flexible 6-membered rings without any double bonds. The substituent orientation angles, α and β, depend on the substituent types and ring sizes and can be predicted accurately from the puckering parameters.

More importantly, we studied the relationship between Cremer−Pople puckering parameters and torsion angles, which facilitated the analysis of the change in endocyclic torsion angles upon pseudo-rotation and other puckering. We have also examined the relationship between endocyclic and exocyclic torsion angles. A knowledge-based ring conformer sampling method based on the puckering preference was proposed, and kernel density estimation (KDE) was used to estimate the puckering preferences. We demonstrated its effectiveness in sampling low-energy small- and medium-sized ring conformations. To progress to larger ring systems, more structural data is necessary for the KDE estimation. Future work should focus on increasing sampling with additional accurate quantum mechanics (QM) energy calculations[57] and developing better density estimation techniques to capture the correlated puckering preferences in large rings. The resulting puckering preferences derived from conformations with QM energies can then be utilized to sample low-energy macrocycle conformations efficiently. Furthermore, our proposed sampling framework can be integrated into other knowledge-based conformer sampling tools, such as Confab[59] and OMEGA,[29] to enhance their sampling performance. We intend to benchmark the performance of our sampling approach with other sampling methods in the future.

We believe that our proposed models and sampling framework are general and readily extensible to larger and more complex ring systems. Further understanding of the conformational preference of cyclic molecules will help accelerate the sampling of low-energy conformers for a wide range of computational modeling applications.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c01144.

> Data and code can be found in GitHub https://github.com/lucianlschan/RING; Cremer−Pople puckering parameters; ring ordering; ring substituent orientation; unique ring families; reconstructing Cartesian coordinates; connection between ring puckering, substituent orientation, and torsion angles; performance metrics; circular correlation; circular distance and variation; distribution of properties; ring conformational preferences; side-chain torsion angles; and Puckering preference in the solid state and gas phase (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Garrett M. Morris − *Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.;* ⊙ orcid.org/0000-0003-1731-8405; Email: morris@stats.ox.ac.uk

Authors

Lucian Chan − *Department of Statistics, University of Oxford, Oxford OX1 3LB, U.K.;* ⊙ orcid.org/0000-0002-1223-0431

Geoffrey R. Hutchison − *Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States; Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States;* ⊙ orcid.org/0000-0002-1757-1980

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.0c01144

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Davies, G. J.; Planas, A.; Rovira, C. Conformational Analyses of the Reaction Coordinate of Glycosidases. *Acc. Chem. Res.* **2012**, *45*, 308−316.

(2) Hancock, R. D.; Martell, A. E. Ligand design for selective complexation of metal ions in aqueous solution. *Chem. Rev.* **1989**, *89*, 1875−1914.

(3) Begel, S.; Puchta, R.; van Eldik, R. Host-guest complexes of calix[4]tubes - prediction of ion selectivity by quantum chemical calculations VI. *J. Mol. Model.* **2014**, *20*, No. 280.

(4) Gong, H.-Y.; Wang, D.-X.; Zheng, Q.-Y.; Wang, M.-X. Highly selective complexation of metal ions by the self-tuning tetraazaca-lixpyridine macrocycles. *Tetrahedron* **2009**, *65*, 87−92.

(5) Marsault, E.; Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. *J. Med. Chem.* **2011**, *54*, 1961−2004.

(6) Giordanetto, F.; Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties. *J. Med. Chem.* **2014**, *57*, 278−295.

(7) Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco, J. A.; Kozakov, D.; Vajda, S.; Whitty, A. How Proteins Bind Macrocycles. *Nat. Chem. Biol.* **2014**, *10*, 723−731.

(8) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The exploration of macrocycles for drug discovery-an underexploited structural class. *Nat. Rev. Drug Discovery* **2008**, *7*, 608−624.

(9) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.

(10) Ionescu, A. R.; Bérces, A.; Zgierski, M. Z.; Whitfield, D. M.; Nukada, T. Conformational Pathways of Saturated Six-Membered Rings. A Static and Dynamical Density Functional Study. *J. Phys. Chem. A* **2005**, *109*, 8096−8105.

(11) Alibay, I.; Bryce, R. A. Ring Puckering Landscapes of Glycosaminoglycan-Related Monosaccharides from Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4729−4741.

(12) Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *J. Am. Chem. Soc.* **2014**, *136*, 1008−1022.

(13) Bocian, D. F.; Pickett, H. M.; Rounds, T. C.; Strauss, H. L. Conformations of cycloheptane. *J. Am. Chem. Soc.* **1975**, *97*, 687−695.

(14) Pakes, P. W.; Rounds, T. C.; Strauss, H. L. Conformations of cyclooctane and some related oxocanes. *J. Phys. Chem. A* **1981**, *85*, 2469−2475.

(15) Anet, F. A. L.; Cheng, A. K. Conformation of cyclohexadecane. *J. Am. Chem. Soc.* **1975**, *97*, 2420−2424.

(16) Al-Jallal, N. A.; Al-Kahtani, A. A.; El-Azhary, A. A. Conformational Study of the Structure of Free 18-Crown-6. *J. Phys. Chem. A* **2005**, *109*, 3694−3703.

(17) El-Azhary, A. A.; Al-Kahtani, A. A. Conformational Study of the Structure of 12-crown-4-Alkali Metal Cation Complexes. *J. Phys. Chem. A* **2005**, *109*, 8041−8048.

(18) Gutsche, C. D.; Bauer, L. J. Calixarenes. 13. The conformational properties of calix[4]arenes, calix[6]arenes, calix[8]arenes, and oxacalixarenes. *J. Am. Chem. Soc.* **1985**, *107*, 6052−6059.

(19) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.

(20) Flores-Ortega, A.; Casanovas, J.; Zanuy, D.; Nussinov, R.; Alemán, C. Conformations of Proline Analogues Having Double Bonds in the Ring. *J. Phys. Chem. B* **2007**, *111*, 5475−5482.

(21) Lyu, S.; Beiranvand, N.; Freindorf, M.; Kraka, E. Interplay of Ring Puckering and Hydrogen Bonding in Deoxyribonucleosides. *J. Phys. Chem. A* **2019**, *123*, 7087−7103.

(22) Appavoo, S. D.; Huh, S.; Diaz, D. B.; Yudin, A. K. Conformational Control of Macrocycles by Remote Structural Modification. *Chem. Rev.* **2019**, *119*, 9724−9752.

(23) Landrum, G. *RDKit: Open-Source Cheminformatics*; BibSonomy, 2016.

(24) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. The Thermodynamics and Molecular Structure of Cyclopentane[1]. *J. Am. Chem. Soc.* **1947**, *69*, 2483−2488.

(25) Hill, A. D.; Reilly, P. J. Puckering Coordinates of Monocyclic Rings by Triangular Decomposition. *J. Chem. Inf. Model.* **2007**, *47*, 1031−1035.

(26) Haasnoot, C. A. G. The conformation of six-membered rings described by puckering coordinates derived from endocyclic torsion angles. *J. Am. Chem. Soc.* **1992**, *114*, 882−887.

(27) Paoloni, L.; Rampino, S.; Barone, V. Potential-Energy Surfaces for Ring-Puckering Motions of Flexible Cyclic Molecules through Cremer−Pople Coordinates: Computation, Analysis, and Fitting. *J. Chem. Theory Comput.* **2019**, *15*, 4280−4294.

(28) Sega, M.; Autieri, E.; Pederiva, F. Pickett angles and Cremer−Pople coordinates as collective variables for the enhanced sampling of six-membered ring conformations. *Mol. Phys.* **2011**, *109*, 141−148.

(29) 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.* **2010**, *50*, 572−584.

(30) Kolodzik, A.; Urbaczek, S.; Rarey, M. Unique Ring Families: A Chemically Meaningful Description of Molecular Ring Topologies. *J. Chem. Inf. Model.* **2012**, *52*, 2013−2021.

(31) Schneider, N.; Sayle, R. A.; Landrum, G. A. Get Your Atoms in Order-An Open-Source Implementation of a Novel and Robust Molecular Canonicalization Algorithm. *J. Chem. Inf. Model.* **2015**, *55*, 2111−2120.

(32) Cremer, D. A General Definition of Ring Substituent Positions. *Isr. J. Chem.* **1980**, *20*, 12−19.

(33) de Leeuw, F. A.; Kampen, P. N. V.; Altona, C.; Díez, E.; Esteban, A. L. Relationships between torsion angles and ring-puckering coordinates: Part III. Application to heterocyclic puckered five-membered rings. *J. Mol. Struct.* **1984**, *125*, 67−88.

(34) Cremer, D. Calculation of puckered rings with analytical gradients. *J. Phys. Chem. A* **1990**, *94*, 5502−5509.

(35) Cole, J. C.; Korb, O.; McCabe, P.; Read, M. G.; Taylor, R. Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *J. Chem. Inf. Model.* **2018**, *58*, 615−629.

(36) Friedrich, N.-O.; Flachsenberg, F.; Meyder, A.; Sommer, K.; Kirchmair, J.; Rarey, M. Conformator: A Novel Method for the Generation of Conformer Ensembles. *J. Chem. Inf. Model.* **2019**, *59*, 731−742.

(37) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database—an open-access collection of crystal structures. *J. Appl. Crystallogr.* **2009**, *42*, 726−729.

(38) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **2012**, *40*, D420−D427.

(39) Sterling, T.; Irwin, J. J. ZINC 15—Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.

(40) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.

(41) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.

(42) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(43) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.

(44) 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.

(45) Chan, L.; Morris, G.; Hutchison, G. *Understanding Conformational Entropy in Small Molecules*; ChemRxiv, 2020.

(46) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Model.* **2020**, *60*, 2044−2058.

(47) Schulz-Gasch, T.; Schärfer, C.; Guba, W.; Rarey, M. TFD: Torsion Fingerprints As a New Measure To Compare Small Molecule Conformations. *J. Chem. Inf. Model.* **2012**, *52*, 1499−1512.

(48) Flachsenberg, F.; Andresen, N.; Rarey, M. RingDecomposer-Lib: An Open-Source Implementation of Unique Ring Families and Other Cycle Bases. *J. Chem. Inf. Model.* **2017**, *57*, 122−126.

(49) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(50) Edman, P. Chemistry of Amino Acids and Peptides. *Annu. Rev. Biochem.* **1959**, *28*, 69−96.

(51) Loiseau, N.; Gomis, J.-M.; Santolini, J.; Delaforge, M.; André, F. Predicting the conformational states of cyclic tetrapeptides. *Biopolymers* **2003**, *69*, 363−385.

(52) Dunbrack, R. L. Rotamer Libraries in the 21st Century. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431−440.

(53) McHugh, S. M.; Rogers, J. R.; Yu, H.; Lin, Y.-S. Insights into How Cyclic Peptides Switch Conformations. *J. Chem. Theory Comput.* **2016**, *12*, 2480−2488.

(54) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. I. Cycloheptane. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *49*, 910−928.

(55) Allen, F. H.; Garner, S. E.; Howard, J. A. K.; Pitchford, N. A. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. III. endo-unsaturated seven-membered rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **1994**, *50*, 395−404.

(56) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. IV. Cyclooctane and related eight-membered rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **1996**, *52*, 882−891.

(57) Rai, B. K.; Sresht, V.; Yang, Q.; Unwalla, R.; Tu, M.; Mathiowetz, A. M.; Bakken, G. A. Comprehensive Assessment of Torsional Strain in Crystal Structures of Small Molecules and Protein−Ligand Complexes using ab Initio Calculations. *J. Chem. Inf. Model.* **2019**, *59*, 4195−4208.

(58) Schrödinger, L. *The PyMOL Molecular Graphics System*, version 1.5, Schrödinger, LLC: New York, 2010.

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

(60) Pordes, R.; Petravick, D.; Kramer, B.; Olson, D.; Livny, M.; Roy, A.; Avery, P.; Blackburn, K.; Wenaus, T.; Würthwein, F. The Open Science Grid. *J. Phys. Conf. Ser.* **2007**, No. 012057.

(61) Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F. In *The Pilot Way to Grid Resources Using glideinWMS*, 2009 WRI World Congress on Computer Science and Information Engineering, 2009; pp 428−432.