Supporting Information: Understanding Ring Puckering in Small Molecules and Cyclic Peptides

Lucian Chan,[†] Geoffrey R. Hutchison,[‡,¶] and Garrett M. Morris[*,†]

†*Department of Statistics, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, U.K.*

‡*Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, PA 15260, U.S.A.*

¶*Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.*

E-mail: morris@stats.ox.ac.uk

Appendix 1

Cremer Pople Puckering Parameters (CP)

Cremer and Pople[S1] proposed a coordinate system to describe the geometry of an N-membered *monocyclic* ring, which requires the specification of an appropriate reference plane. The mathematical details are given below.

Let the position of each atom, j, in the N-membered puckered ring be specified by the Cartesian coordinates (X_j, Y_j, Z_j); and let $\mathbf{R_j}$ be the corresponding position vector of ring atom j, with the origin, (0, 0, 0) as the geometrical center of the ring, such that it satisfies Equation (S1):

$$\sum_{j=1}^{N} \mathbf{R_j} = \mathbf{0} \tag{S1}$$

From this constraint, it follows that $\sum_{j=1}^{N} z_j = 0$, and effectively suppresses translation of the planar reference. Two additional constraints, Equations (S2) and (S3), are imposed to suppress overall rotation of the planar reference about the x- and y-axes:

$$\sum_{j=1}^{N} z_j \cos\left(\frac{2\pi(j-1)}{N}\right) = 0 \tag{S2}$$

$$\sum_{j=1}^{N} z_j \sin\left(\frac{2\pi(j-1)}{N}\right) = 0 \tag{S3}$$

The orientation of the mean plane can now be determined by two vectors, $\mathbf{R}'$ and $\mathbf{R}''$, as in Equations (S4) and (S5), where $\mathbf{n}$ is the unit normal vector to the mean plane, Equation (S6); note that the positive direction of $\mathbf{n}$ defines the "top" side of the ring.

$$\mathbf{R}' = \sum_{j=1}^{N} R_j \sin\left(\frac{2\pi(j-1)}{N}\right) \tag{S4}$$

$$\mathbf{R}'' = \sum_{j=1}^{N} R_j \cos\left(\frac{2\pi(j-1)}{N}\right) \tag{S5}$$

$$\mathbf{n} = \frac{\mathbf{R}' \times \mathbf{R}''}{|\mathbf{R}' \times \mathbf{R}''|} \tag{S6}$$

Using the ring atom position vector and the unit normal vector, $\mathbf{n}$, defined by $\mathbf{R}'$ and $\mathbf{R}''$, we can compute the full set of displacements from the mean plane, z_j, for ring atoms $j = 1,\ ...,\ N$, using the scalar products in Equation (S7); this will also satisfy Equations (S2) and (S3) automatically:

$$z_j = \mathbf{R_j} \cdot \mathbf{n} \tag{S7}$$

The general ring-puckering coordinates for an N-membered ring are calculated as follows:

If N is odd and $N >3$, we define q_m and ϕ_m using:

$$q_m \cos\phi_m = \left|\sqrt{\frac{2}{N}}\right| \sum_{j=1}^{N} z_j \cos\left(\frac{2\pi m(j-1)}{N}\right) \tag{S8}$$

$$q_m \sin\phi_m = -\left|\sqrt{\frac{2}{N}}\right| \sum_{j=1}^{N} z_j \sin\left(\frac{2\pi m(j-1)}{N}\right) \tag{S9}$$

These formulae apply for $m = 2,\ 3,\ ...,\ \frac{(N-1)}{2}$. They represent a set of puckering coordinates with zero or positive amplitudes, q_m, ($q_m \geq 0$), and phase angles, ϕ_m ($-\pi \leq \phi_m < \pi$).

If N is even, the coordinates in Equation (S8) and (S9) apply up to $m = \frac{N}{2} - 1$, but there is an additional puckering coordinate, as shown in Equation (S10); note that the amplitude $q_{\frac{N}{2}}$ can take either sign:

$$q_{\frac{N}{2}} = \sqrt{\frac{1}{N}} \sum_{j=1}^{N} z_j (-1)^{j-1} \tag{S10}$$

Ring Ordering

The Cremer-Pople ring puckering parameters provide a quantitative means to describe the atomic coordinates of a ring. We should note that the representation is atom-order dependent. The choice of the first atom and the atom numbering order, *i.e.* clockwise or anti-clockwise, will affect the outcome. To overcome this ambiguity, we introduced a new atom numbering scheme that is based on the bond order, connectivity (number of substituents attached to the ring atoms), and the element types in the ring. The first atom is chosen with the highest bond order, *i.e.* triple > double > aromatic > single. If the bond order of all the ring bonds is the same, such as in cyclohexane, then atomic connectivity is used to determined the first atom. Atoms with exocyclic double bonds have the highest priority, followed by atoms with two non-hydrogen substituents, followed by one non-hydrogen substituent, and finally no substituents. If the first atom cannot be determined by the rules above, then the atom with the minimum atomic number is assigned as the first atom. Otherwise, the first atom is assigned randomly when the ring is symmetric, such as in cyclohexane.

The direction of numbering can be determined by the sum of the ring bonds' bond orders, ring atom connectivity, and ring atom element types in a ring path. The ring path takes $\frac{N+2}{2}$ atoms when N is even, and $\frac{N+1}{2}$ atoms when N is odd. For any N-membered ring, the path with the largest sum of bond orders is chosen. Similarly, the path with the largest connectivity sum is selected when there are multiple paths with same bond orders (an example is shown in Figure S1). If multiple possibilities exist, the path with the minimum sum of the atomic numbers from the first atom to atom $\frac{N+2}{2}$ when N is even (or atom $\frac{N+1}{2}$ when N is odd) is chosen.

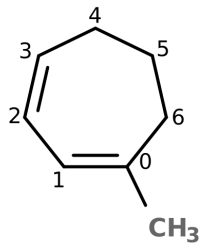


Figure S1: Atom ordering for 1-Methylcyclohepta-1,3-diene. There are two possible orderings, including: (0,1,2,3,4,5,6) and (3,2,1,0,6,5,4), based on the bond order criterion. Since atom 0 has one non-hydrogen (methyl) substituents, it has a higher connectivity than atom 3. Thus, the final atom ordering is:(0,1,2,3,4,5,6).

For cyclic peptides, we introduced an additional criterion for ring atom numbering. We take the volume of the amino acid side chain of the amino acid into account, and giving highest priority to the bulkiest side chain, trypotophan, and so on down to glycine. The amino acid ranks are listed in Table S1.

Table S1: Amino Acid Volume Ranking.

Amino Acid	Volume (Å^3)	Rank
Tryptophan	227.8	1
Tyrosine	193.6	2
Phenylalanine	189.9	3
Arginine	173.4	4
Lysine	168.6	5
Isoleucine	166.7	6
Leucine	166.7	6
Methionine	162.9	8
Histidine	153.2	9
Glutamine	143.8	10
Valine	140.0	11
Glutamic acid	138.4	12
Threonine	116.1	13
Asparagine	114.1	14
Proline	112.7	15
Aspartic acid	111.1	16
Cysteine	108.5	17
Serine	89.0	18
Alanine	88.6	19
Glycine	60.1	20

Ring Substituent Orientation

Ring substituents play an important role in determining ring geometries. To quantify their role, we used a coordinate system to describe each substituent's orientation, which is complementary to Cremer-Pople puckering parameters.[S2] The mathematical details are given below.

Let $\mathbf{n}$ be the unit vector perpendicular to the mean plane (as defined in Cremer-Pople puckering parameters); let $\mathbf{s}_j$ be the unit vector pointing from a ring atom, j, to the corresponding substituent, S; let $\mathbf{u}$ be the vector that points from the geometrical center to the projection of the position of the ring atom onto the mean plane; and let $\mathbf{v}$ be the vector perpendicular to both $\mathbf{n}$ and $\mathbf{u}$. The substituent orientation can now be described by two angles, α and β:

α ranges from 0 to π radians, while β ranges from $-\pi$ to π radians. The ring substituent angles, α and β, are defined by Equations (S11), (S12) and (S13).

$$\cos\alpha = \mathbf{s}_j \cdot \mathbf{n} \tag{S11}$$

$$\mathbf{s}_j \cdot \mathbf{u}_j = \sin\alpha\cos\beta \tag{S12}$$

$$\mathbf{s}_j \cdot \mathbf{v}_j = -\sin\alpha\cos\beta \tag{S13}$$

Unique Ring Familes (URFs)

To study the puckering preferences of complex rings system such as fused rings and spiro rings, we apply the concept of unique ring families[S3] (URFs) to decompose complex ring systems into multiple sub-groups. We then apply a standardized numbering scheme in all sub-groups and calculate their puckering parameters. This decomposition is also used in ring reconstructions, as described in the next section.

The calculation of URFs comprises two parts: (i) calculation of Relevant Cycles (RCs), where the RCs are defined as the union of all minimum cycle bases; and (ii) pairing of RCs if they are URF-pair-related (see Definition 1).

Definition 1. *Let C_1 and C_2 be two RCs in a molecular graph, G; then C_1 and C_2 are URF-pair related if and only if all of the following conditions hold:*

1. *$|C_1| = |C_2|$, i.e. the number atoms in each ring is the same;*
2. *$E(C_1) \bigcap E(C_2)$, i.e. the two rings share one or more bonds; and*
3. *There exists a set, S, of strictly smaller rings, c, in G such that $C_1 \bigoplus (\bigoplus_{c\in S} c) = C_2$.*

Reconstructing Cartesian Coordinates from Cremer-Pople Puckering Parameters

Cremer-Pople puckering parameters are uniquely defined and can be easily calculated from the Cartesian coordinates of the atoms in a N-membered ring. On the other hand, the

Cartesian coordinates of any N-membered rings can be derived from its $N-3$ puckering parameters and additional $2N-3$ internal coordinates that describe the planar reference ring. The calculation comprises the following steps:

1. Calculation of z evaluations;
2. Projection of bond lengths and bond angles on the mean plane;
3. Ring partition; and
4. Ring partition coordinate calculations.

Step 1: Calculation of z elevations

Cremer-Pople puckering parameters are required for this calculation, and the inversion formulae for an N-membered ring are given below:

$$z_j = \left|\sqrt{\frac{2}{N}}\right| \sum_{m=2}^{\frac{N-1}{2}} q_m \cos\left(\phi_m + \frac{2\pi m(j-1)}{N}\right), \qquad \text{if } N \text{ is odd} \tag{S14}$$

$$z_j = \left|\sqrt{\frac{2}{N}}\right| \sum_{m=2}^{\frac{N}{2}-1} q_m \cos\left(\phi_m + \frac{2\pi m(j-1)}{N}\right) + \frac{1}{N} q_{\frac{N}{2}} (-1)^{j-1}, \qquad \text{if } N \text{ is even} \tag{S15}$$

for ring atoms $j = 1, ..., N$.

Step 2: Projection of bond lengths and bond angles

Once the z_j coordinates have been determined, the bond lengths, r', and bond angles, β', of the planar reference ring can be computed by projecting the N bond lengths, r, and $N-3$ bond angles, β, of the puckered ring onto the ring's mean plane. The N projected bond lengths, r, and $N-3$ projected bond angles, β', suffice to determine the (x, y) coordinates of the projected atoms. The initial bond lengths and bond angles are listed in Tables S2 and S3, respectively. Note that the reference bond lengths and bond angles will vary with ring size. For ring atoms i, j, and k:

$$r'_{ij} = \left| \sqrt{r_{ij}^2 - (z_i - z_j)^2} \right| \tag{S16}$$

$$\cos(\beta'_{ijk}) = \frac{(z_k - z_i)^2 - (z_k - z_i)^2 - (z_k - z_i)^2 + 2r_{ij}r_{jk}\cos(\beta_{ijk})}{2r'_{ij}r'_{jk}} \tag{S17}$$

Table S2: Reference Bond Lengths. Aromatic atoms are represented by lower case.

Bond	Ring Size, N	Bond Length (Å)
C—C	5	1.535
C=C	5	1.339
c—c	5	1.403
C—C	6	1.531
C=C	6	1.336
c—c	6	1.388
C—C	7	1.530
C=C	7	1.339
c—c	7	1.390

Table S3: Reference Bond Angles. Aromatic atoms are represented by lower case.

Bond Angle	Ring Size, N	Bond Angle (rad)
C—C—C	5	1.816
C=C—C	5	1.904
c—c—C	5	1.859
C—C—C	6	1.931
C=C—C	6	2.132
c—c—C	6	2.097
C—C—C	7	1.979
C=C—C	7	2.188
c—c—C	7	2.256

Step 3: Ring partition

To calculate the (x, y) coordinates of the planar reference ring, we partition the ring into three segments, S_1, S_2 and S_3, by inscribing a triangle inside the ring. The number of atoms, bond length, r, and number of bond angles, β, in each fragment are listed in Table S4. This partitioning keeps the error progression due to coupling of the ring parameters at a minimum.

Table S4: Partitioning of a Projected Planar Ring into Segments S_1, S_2 and S_3. The number of atoms, bonds, and angles in each of the three segments are given for N-membered rings of varying size.

	# Atoms in Segment			# Bonds in Segment			# Angles in Segment		
N	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
5	3	2	3	2	1	2	1	0	1
6	3	3	3	2	2	2	1	1	1
7	3	4	3	2	3	2	1	2	1
8	4	3	4	3	2	3	2	1	2
9	4	4	4	3	3	3	2	2	2
10	4	5	4	3	4	3	2	3	2
11	5	4	5	4	3	4	3	2	3
12	5	5	5	4	4	4	3	3	3
13	5	6	5	4	5	4	3	4	3
14	6	5	6	5	4	5	4	3	4
15	6	6	6	5	5	5	4	4	4
16	6	7	6	5	6	5	4	5	4

Step 4: Ring partition coordinate calculations.

The (x, y) coordinates of the atoms of each segment can be calculated from the r' and β' values. In the next step, the coordinates of the vertices of the inscribed triangle are determined by lengths of the sides of the triangle, namely R_1, R_2, and R_3, which are equivalent to the distance between the terminal atoms of each segment. One point of this triangle is placed at the origin, and second point on the positive x-axis. The coordinates of the remaining vertex of the triangle can be calculated easily using R_1, R_2, and R_3. In the remaining steps, the three segments are rotated such that the terminal atoms coincide with appropriate segments. Finally, the planar ring is translated such that the geometric center coincides with the origin, thus yielding the final set of Cartesian coordinates.

The procedure outlined can be applied to any N-membered ring with or without symmetry. For complex rings, this procedure is applied to all URFs in a molecule. Note that this procedure is sensitive to the puckering parameters, bond lengths and bond angles in the planar reference ring. Poor parameter choice will lead to numerical error.

Connection between Ring Puckering, Substituent Orientation and Torsion Angles

It is well known that the substituent orientation changes upon puckering. Pseudo-rotation, a specific type of ring puckering, is of interest. To understand the change of substituent orientation upon pseudo-rotation, we propose two models, 1 and 2, to predict the orientation angles, α and β, of a substituent. Both models share the same functional form (Equation S18), but the estimated parameters, A_i, $B_{i,m}$, $C_{i,m}$, D_i, $E_{i,m}$, $F_{i,m}$, and G_i will be different. Note that α_i and β_i are the orientation angles of the substituents attached to ring atom, i, and N is the number of atoms in the ring:

Model 1 and 2:

$$\begin{aligned}\alpha_i, \beta_i = A_i + \sum_{m=2}^{M} B_{i,m} q_m \cos(\phi_m + \frac{2\pi m(i-1)}{N}) + \sum_{m=2}^{M} C_{i,m} q_m \sin(\phi_m + \frac{2\pi m(i-1)}{N}) + \\ D_i \mathbb{1}_{N,2(M+1)} (-1)^{i-1} q_{\frac{N}{2}} + \\ \sum_{m}^{M} E_{i,m} q_m^2 \cos(2(\phi_m + \frac{2\pi m(i-1)}{N})) + \sum_{m}^{M} F_{i,m} q_m^2 \sin(2(\phi_m + \frac{2\pi m(i-1)}{N})) + \\ G_i \mathbb{1}_{N,2(M+1)} (-1)^{i-1} q_{\frac{N}{2}}^2\end{aligned} \tag{S18}$$

where q_m, ϕ_m are the puckering amplitudes and phase angles respectively. $M = \frac{N-1}{2}$ when N is odd, otherwise $M = \frac{N}{2} - 1$. $\mathbb{1}_{N,2(M+1)}$ is an indicator function, with 1 when $N = 2(M+1)$.

Torsion angles are an alternative way to measure or define ring geometries. They are widely used in conformational analysis of small rings. Inspired by the functional forms used in previous works,[S4] the Model 3, Equation (S19), can be used to map Cremer-Pople puckering parameters to endocyclic torsion angles for general N-membered rings. Note that the functional forms used in de Leeuw et al. can be recovered by applying trigonometric identities and setting $N = 5$.

Model 3:

$$\theta_i^{endo} = H_i + \sum_{m=2}^{M} J_{i,m} q_m \cos(\phi_m + \frac{2\pi m(i-1)}{N}) + \sum K_{i,m} q_m \sin(\phi_m + \frac{2\pi m(i-1)}{N}) + L_i \mathbb{1}_{N,2(M+1)} (-1)^{i-1} q_{\frac{N}{2}} \tag{S19}$$

where $H_i, J_{i,m}, K_{i,m}, L_i$ are the model parameters. q_m, ϕ_m, and $\mathbb{1}_{N,2(M+1)}$ are defined as above. θ_i^{endo} is the endocyclic torsion angles defined by four ring atoms ($i-1$, i, $i+1$, $i+2$).

Furthermore, a substituent's orientation can also be expressed in terms of torsion angles and bond angles. Here, we focus on the relationship between endocyclic and exocyclic torsion angles, and a linear model, Model 4, Equation (S20), is proposed:

Model 4:

$$\theta_i^{exo} = P_i + Q_i \theta_i^{endo} \tag{S20}$$

where P_i and Q_i are model parameters. Endocyclic torsion angles are defined by the four ring atoms ($i-1$, i, $i+1$, $i+2$), θ_i^{endo}; while θ_i^{exo} is the exocyclic torsion angle defined by the substituent atom, s_i attached to ring atom i, and the three ring atoms (i, $i+1$, $i+2$), so the exocyclic torsion is defined by (s_i, i, $i+1$, $i+2$).

For all models, a random sample of 50% of the data was used to estimate the model parameters, and the rest was used to assess the performance of the models. The associated model parameters were estimated by the least-squares method for all models. The performance metrics are discussed below.

Performance Metrics

Two metrics were used to assess the quality of the generated conformations, namely heavy atom root-mean-square deviation (RMSD) and Torsion Fingerprint Deviation (TFD).[S5] The lowest energy conformation obtained from CREST was used as reference conformation in

each case. RMSD is defined as follows:

$$\text{RMSD} = \left| \sqrt{\frac{1}{N_{\text{atoms}}} \sum_{i=1}^{N_{\text{atoms}}} (r_i - r_{i,\text{ref}})^2} \right| \tag{S21}$$

where N_{atoms} is the number of heavy atoms in the molecule, r_i is the position of atom i, and $r_{i,\text{ref}}$ is the position of the corresponding atom i in the lowest energy CREST reference structure. Symmetry is taken into account for the calculation. The implementations of RMSD and TFD calculation in RDKit were used.

Three metrics: circular correlation coefficient R^2_{circ}, mean angular error (MAE), and standard deviation angular error (circular distance between predicted and actual angles), were used to assess the performance of our proposed models.

Circular Correlation

Since (torsion) angles are the same when adding or subtracting integer multiples of one whole rotation, we need to use a metric that takes this rotational equivalence into account. Let θ^a and θ^b be two circular variables; μ_a and μ_b be the mean directions of θ^a and θ^b, respectively. The mean direction of a circular variable can be computed using Equations (S24) and (S26) below. The circular correlation coefficient, ρ_c, is defined as follows:

$$\rho_c(\theta^a, \theta^b) = \frac{\mathbb{E}(\sin(\theta^a - \mu_a)\sin(\theta^b - \mu_b))}{\sqrt{\text{Var}(\sin(\theta^a - \mu_a))\text{Var}(\sin(\theta^b - \mu_b))}} \tag{S22}$$

The circular correlation coefficient varies from -1 to 1, and is 0 when the two circular variables are independent. We report ρ_c^2, *i.e.*, the square of the circular correlation coefficient, and henceforth refer to this as "R^2_{circ}".

Circular Distance and Variation

The standard Euclidean distance is not applicable to angular data, due to its circular nature. To measure the distance between two angular variables, θ^a and θ^b, we must compute the minimum arc length, $d_0(\theta^a, \theta^b)$, between them:

$$d_0(\theta^a, \theta^b) = \min(\theta^a - \theta^b,\ 2\pi - (\theta^a - \theta^b)) \tag{S23}$$

Similarly, for a set of angular variables, θ_i, the mean, $\frac{1}{N}\sum_{i=1}^{N}\theta_i$, is not well defined, as it depends on where the circle is cut (here, N is the number of angular variables, not the number of ring atoms as before). We compute the mean direction of a set of angular variables as follows:

$$(\bar{C}, \bar{S}) = \left(\frac{\sum_{i=1}^{N} \cos\theta_d}{N},\ \frac{\sum_{i=1}^{N} \sin\theta_d}{N} \right) \tag{S24}$$

The length of the resultant mean, $\bar{R}$, is given by:

$$\bar{R} = \left| \sqrt{(\bar{C} + \bar{S})} \right| \tag{S25}$$

The direction of the mean, $\bar{\theta}$, is given by:

$$\bar{\theta} = \begin{cases} \tan^{-1}\frac{\bar{S}}{\bar{C}} & \text{if } \bar{C} > 0 \\ \tan^{-1}\frac{\bar{S}}{\bar{C}} + \pi & \text{if } \bar{C} < 0 \end{cases} \tag{S26}$$

The circular standard deviation, $\sigma(\theta)$, can be computed as follows:

$$\sigma(\theta) = \left| \sqrt{-2\log(\bar{R})} \right| \tag{S27}$$

Appendix 2: Relationship between Puckering Parameters, Substituent Orientation, Endocyclic and Exocyclic Torsion Angles

The fitted parameters of the models introduced in Appendix 1 are listed below. Here, we only report the parameters of the conformational clusters that appeared in our analysis.

Table S5: Model parameters for predicting the α substituent orientation angle of carbonyl functional group at a given position. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation (s.d.) of angular error.

Substituent	Ring Size (Cluster)	Model Parameters (Pos, A, B, C, D, E, F, G)	Performance (R^2_{circ}, MAE, s.d.)
Carbonyl	5 (Envelope)	(1, 1.572, -1.185, -0.016, -0.001, -0.051) (2, 1.576, -1.217, -0.000, 0.045, -0.020) (3, 1.572, -1.249, 0.079, 0.013, 0.114)	(0.990, 0.026, 0.028) (0.986, 0.023, 0.028) (0.994, 0.022, 0.016)
Carbonyl	6 (Chair)	(1, 0.534, -1.343, 0.084, 3.302, -0.071, -0.006, -4.443), (2, 1.473, -1.340, -0.059, 0.995, 0.318, 0.005, 0.417)	(0.974, 0.032, 0.033), (0.958, 0.029, 0.032)
Carbonyl	6 (Chair)	(1, 0.880, -1.359, 0.014, -3.185, 0.129, 0.165, -1.568), (2, 1.603, -1.264, 0.075, 0.725, -0.002, 0.244, -0.694)	(0.974, 0.028, 0.026), (0.949, 0.034, 0.037)
Carbonyl	6 (Boat)	(1, 1.577, -1.104, 0.007, -1.270, 0.015, 0.015, -1.107) (2, 1.575, -1.104, 0.004, 1.427, 0.009, 0.018, -0.489)	(0.997, 0.035, 0.031), (0.997, 0.032, 0.035),
Carbonyl	12 (CCCC-DDDD, Sub-cluster 1)	(1, -1.840, -0.376, -0.708, -0.225, -1.071, 0.056, -0.324, -6.098, -0.156, -0.508, -0.061, 0.411, 1.803, 0.901, 0.043, 0.182, -1.001, -0.4201, 0.481) (4, 2.353, -0.379, -0.730, 2.795, -1.124, 0.061, -0.345, -0.959, -0.356, 0.451, 0.063, 0.422, 0.995, 3.512, 0.026, 0.160, 0.933, 0.436, 3.163) (7, 3.2801, -0.3783, -0.7161, 2.929, -1.1372, 0.0725, -0.3533, 0.4037, -0.2139, -0.3711, 0.0275, 0.3781, 0.6678, -0.3031, 0.0196, 0.1622, 1.2203, 1.7649, 1.1696) (10, 1.291, -0.347, -0.703, -1.845, -1.192, 0.025, -0.367, 0.538, -0.244, 0.551, 0.070, 0.289, -0.370, 1.066, 0.013, 0.288, -0.079, 1.466, 2.157)	(0.995, 0.012, 0.009) (0.994, 0.013, 0.010) (0.994, 0.013, 0.010) (0.994, 0.013, 0.010)

Table S5: Model parameters for predicting the α substituent orientation angle of carbonyl functional group at a given position. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation (s.d.) of angular error.

Substituent	Ring Size (Cluster)	Model Parameters (Pos, A, B, C, D, E, F, G)	Performance (R^2_{circ}, MAE, s.d.)
Carbonyl	12 (CCCC-DDDD, Sub-cluster 2)	(1, 0.519, -0.375, -0.586, -3.332, -0.861, 0.037, -0.113, -0.823, -0.028, -0.770, 0.017, -0.213, -0.267, -0.017, -0.057, 0.632, -1.086, 0.101, 1.713) (4,1.550, -0.376, -0.453, -0.953, -0.988, 0.038, -0.144, -0.422, -0.095, 0.701, 0.011, 0.187, 0.104, -0.339, -0.045, 0.619, -0.216, 2.095, -2.561) (7, 0.838, -0.380, -0.534, 0.988, -0.895,,0.044, -0.169, -3.195, -0.011, -0.725, 0.021, 1.069, 1.567, -0.493, -0.035, -1.400, -0.209, 4.376, -0.236) (10, 0.079, -0.377, -0.457, -2.470, -0.951, 0.044, -0.072, -2.296, 0.008, ,0.801, 0.016, 0.797,0.442, -0.342, -0.049, -0.985, -1.147, 2.500, 4.496)	(0.999, 0.010, 0.010) (0.999, 0.011, 0.010) (0.999, 0.011, 0.010) (1.000, 0.010, 0.009)

Table S6: Model parameters for predicting β orientation angle of carbonyl group at a given position. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation (s.d.) of angular error.

Substituents	Ring Size (Cluster)	Model Parameters (Pos, A, B, C, D, E, F, G)	Performance (R^2_{circ}, MAE, s.d.)
Carbonyl	5 (Envelope)	(1, 0.001, -0.000, 0.002, 0.011, 0.114) (2, 0.001, -0.002, 0.002, 0.010, 0.118) (3, 0.012, -0.23, 0.025, 0.128, 0.037)	(0.410, 0.017, 0.15) (0.620, 0.014, 0.012) (0.337, 0.016, 0.012)
Carbonyl	6 (Chair)	(1, 0.471, 0.038, 0.240, -1.749, 0.064, 0.154, 1.614), (2, -0.278, -0.001, -0.175, 0.929, -0.001, 0.0743, -0.778)	(0.514, 0.021, 0.022), (0.570, 0.015, 0.018)
Carbonyl	6 (Chair)	(1, -0.490, -0.072, -0.276, -1.919, 0.175, 0.159, -1.844), (2, -0.405, 0.019, 0.245, -1.427, 0.018, 0.216, -1.243)	(0.514, 0.021, 0.022), (0.570, 0.0315, 0.018)
Carbonyl	6 (Boat)	(1, 0.002, -0.007, 0.008, -0.102, 0.015, 0.211, -1.237) (2, -0.003, 0.004, -0.003, 0.008, -0.010, 0.213, 0.429)	(0.929, 0.028, 0.025), (0.934, 0.027, 0.025),

Table S6: Model parameters for predicting β orientation angle of carbonyl group at a given position. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation (s.d.) of angular error.

Substituents	Ring Size (Cluster)	Model Parameters (Pos, A, B, C, D, E, F, G)	Performance (R^2_{circ}, MAE, s.d.)
Carbonyl	12 (CCCC-DDDD, Sub-cluster 1)	(1, 21.764, 0.302, -0.3618, 8.760, -0.883, -1.148, -0.176, 32.280, 0.492, -5.984, 0.267, -5.709, -7.149, -24.586, 0.0054, -2.873, 9.648, -20.494, 1.896) (4, -19.002, 0.341, -0.043, -4.879, -0.356, -1.190, 0.021, -30.778, 1.332, 6.009, 0.052, -5.867, 7.729, -21.298, 0.083 ,-2.792, -7.935, -20.719, -5.047) (7, 22.466, 0.405, -0.336, 4.221, 0.835, -1.300, 0.346, 37.008, 0.456, -6.296, 0.166, -6.015, -9.446, -24.509, 0.092, -2.601, 9.392 -20.931, 0.889) (10, -12.859, 0.337, -0.527, -1.170, -0.230, -1.192, 0.054, -22.646, -0.112, 6.009, 0.072, -5.897, 6.250, -21.932, 0.079, -2.597, -5.140, -19.561, -3.109)	(0.908, 0.067, 0.051), (0.914, 0.066, 0.048), (0.912, 0.065, 0.049), (0.909, 0.068, 0.048)
Carbonyl	12 (CCCC-DDDD, Sub-cluster 2)	(1, -5.850, -0.188, -1.256, -19.314, 0.870, -0.393, -0.101, 1.369, 1.258, 0.132, 0.155, -3.474, -4.628, 2.818, 0.028, 5.924, -3.474, -4.628, 19.583) (4, -9.498, -0.134, -0.689, -8.553, 0.414, -0.423, -0.245, -13.320, 0.997, 0.024, 0.127, 0.373, 2.656, -4.144, 0.110, -1.111, -5.818, 8.279, 17.277) (7, -3.751, -0.167,-0.916, -15.416, 0.366, -0.395, -0.127, 4.390, 1.110, 0.290, 0.215, 0.602, -4.754, 1.873, 0.166, -1.458, -3.511, -1.313, 20.958) (10, -7.111, -0.126, -0.985, -5.345, 0.519, -0.430, -0.055, -11.107, 1.507, -0.035, 0.141, -1.687, 3.053, -1.491, 0.081, 1.585, -4.188, -3.969, 24.305)	(0.994, 0.041, 0.044), (0.994, 0.043, 0.042), (0.993, 0.046, 0.046), (0.995, 0.040, 0.044)

Table S7: Model parameters for predicting endocyclic torsion angles. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation (s.d.) of angular error.

Ring Size	Cluster	Model Parameters	Performance (R^2_{circ}, MAE, s.d.)
5	Cluster 1 (Envelope)	(0, 0.000, -1.694, -0.543) , (1 , 0.000, 1.035, 1.433), (2, 0.000, 0.001, -1.803), (3, 0.000, -1.070, 1.483), (4, 0.000, 1.710, -0.575)	(0.999, 0.012, 0.018), (0.999, 0.016, 0.023), (0.999, 0.016, 0.023), (0.999, 0.027, 0.033), (0.999, 0.024, 0.029)
6	Cluster 1 (Chair)	(0, 0.141, -1.155, -0.642, -1.454) , (1, -0.116, -1.111, -0.628, 1.483), (2, 0.160, -1.112, -0.625, -1.412), (3, -0.249, -1.085, -0.632, 1.285), (4, 0.327, -1.029, -0.674, -1.174), (5, -0.268, -1.098, -0.735, 1.264)	(0.971, 0.019, 0.018), (0.964, 0.021, 0.018), (0.953, 0.021, 0.019), (0.938, 0.021, 0.024), (0.889, 0.036, 0.027), (0.917, 0.033, 0.027)
6	Cluster 2 (Chair)	(0, -0.160, -1.155, -0.646, -1.419) , (1, 0.114, -1.084, -0.626, 1.486), (2, -0.174, -1.120, -0.638, -1.385), (3, 0.251, -1.072, -0.617, 1.2806), (4, -0.314, -1.026, -0.659, -1.195), (5, 0.279, -1.067, -0.720, 1.240)	(0.974, 0.018, 0.016), (0.968, 0.021, 0.017), (0.958, 0.022, 0.019), (0.933, 0.022, 0.025), (0.902, 0.034, 0.028), (0.926, 0.031, 0.026)
6	Cluster 3 (Boat)	(0, -0.002, -1.251, -0.707, -1.585) (1, -0.003, -1.251, -0.715, 1.616) (2, 0.001, -1.24, -0.704, -1.658) (3, -0.003, -1.246, -0.741, 1.641) (4, -0.003, -1.299, -0.759, -1.541) (5, 0.003, -1.302, -0.702, 1.549)	(0.999, 0.025, 0.028), (0.999, 0.030, 0.022), (0.999, 0.030, 0.027), (0.999, 0.028, 0.034), (0.999, 0.041, 0.033), (0.998, 0.038, 0.037),

Table S7: Model parameters for predicting endocyclic torsion angles. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error (MAE), and standard deviation of angular error.

Ring Size	Cluster	Model Parameters	Performance (R^2_{circ}, MAE, s.d.)
12	Sub-cluster 1 (CCCC-DDDD)	(1, 1.654 , 0.168, -1.333, -0.834, -0.457, -1.491, 0.803, 0.240, -0.310, -8.318),	(0.884, 0.124, 0.094),
		(2, -1.429, -1.161, 0.731, -0.595, -0.872, -0.211 , -1.001, -1.585, -1.998, 7.736),	(0.824, 0.126, 0.089),
		(4, 2.323, 0.127, -0.773, -0.975, -1.130, -1.390, 0.988, 0.886, 1.102, 8.203),	(0.876, 0.137, 0.096),
		(5, -2.195, -1.191, 0.978, -0.023, -0.437, -0.202, -1.543, -1.046, -0.402, -8.151),	(0.796, 0.132, 0.090),
		(7, 1.452, 0.258, -1.043, -0.790, 1.035, -1.636, 1.637, 0.031, 1.319, -8.982),	(0.875, 0.135, 0.099),
		(8, -1.401, -1.272, 1.519, -0.618, 1.064 , -0.183, -1.375, -1.642, -0.812, 8.332),	(0, 779, 0.129, 0.095),
		(10, 2.5168, 0.088, -1.469, -0.924, 1.127, -1.362, 1.438, 1.020, 0.411, 8.112),	(0.883, 0.133, 0.093),
		(11, -2.380, -1.100, 1.382, 0.099, 0.931, -0.195, -0.863, -1.102, -2.154, -7.874)	(0.774, 0.126, 0.090),
12	Sub-cluster 2 (CCCC-DDDD)	(0, 0.985, -0.187, -0.850, -1.433, -1.600, -0.564, -0.205, 0.050, 0.317, -0.656),	(0.997, 0.037, 0.042),
		(1, -1.384, -0.385, 0.214, -0.664, -0.888, -0.456, -1.228, -1.861, -1.2400, 0.881),	(0.993, 0.045, 0.050),
		(3, 1.106, -0.178, -1.112, -1.577, -1.837, -0.532 , -1.225 , 0.298, -1.126, 1.296),	(0.997, 0.047, 0.043),
		(4, -1.214, -0.416, -1.187, -0.738, -1.785, -0.470, -1.270, -1.461, -2.301, -2.131),	(0.989, 0.060, 0.057),
		(6, 0.971,-0.160, -0.911, -1.441, -1.631, -0.568, 0.030, 0.037, -0.552, -1.201),	(0.997, 0.042, 0.039),
		(7, -1.234, -0.384, 0.519, -0.796, -1.232, -0.443, -1.094, -1.940, -1.385, 0.822),	(0.992, 0.051, 0.051),
		(9, 1.095, -0.170 , -1.380, -1.731, -1.315, -0.568, -0.158, 0.401, 0.033, 0.538),	(0.997, 0.040, 0.041)
		(10, -1.413, -0.385, 0.095, -0.547, -1.553, -0.426, -0.891, -1.304 , -1.357, -0.557)	(0.993, 0.048,0.048)

Table S8: Model parameters for predicting substituent exocyclic torsion angles. Each model's performance is given in terms of circular correlation, R^2_{circ}, mean angular error, MAE, and standard deviation (s.d.) of angular error.

Substituent(s)	Model Parameters (Intercept, Slope)	(R^2_{circ}, MAE, s.d.)
Carbonyl (C=O)	$(-\pi,\ 1)$	(0.998,0.040,0.039)
Methyl (CH3)	$(-\frac{2\pi}{3},\ 1)$, $(-\frac{5\pi}{6},\ 6)$, $(-\pi,\ 1)$, $(\frac{5\pi}{6},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.997,0.052,0.039),(0.994,0.061,0.041),(0.993,0.029,0.028), (0.998,0.056,0.039), (0.998,0.047,0.038)
Hydroxyl (OH)	$(-\frac{2\pi}{3},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.998,0.054,0.038),(0.998,0.057,0.040)
Alkoxy (-O-)	$(-\frac{2\pi}{3},\ 1)$, $(-\pi,\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.998,0.044,0.039),(0.980,0.038,0.037),(0.998,0.048,0.041)
Bulky carbon (-CH0, CH1, CH2)	$(-\frac{2\pi}{3},\ 1)$, $(-\frac{5\pi}{6},\ 6)$, $(\frac{5\pi}{6},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.996, 0.064, 0.052),(0.972, 0.063, 0.036), (0.975, 0.082, 0.049),(0.996, 0.068, 0.054)
Halogen (-F)	$(-\frac{2\pi}{3},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.998,0.046,0.033), (0.998,0.048,0.036)
Halogen (-Cl)	$(-\frac{2\pi}{3},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.996,0.060,0.048), (0.997,0.054,0.043)
Halogen (-Br)	$(-\frac{2\pi}{3},\ 1)$, $(\frac{2\pi}{3},\ 1)$	(0.998,0.057,0.045), (0.998,0.052,0.052)

Appendix 3: Datasets and Reference Structure Calculations

We used three datasets we used in our analysis: (i) 63,814 small organic molecules from Crystallography Open Database (COD); (ii) 67,009 small organic molecules from ZINC; (iii) 8,661 cyclic tetrapeptides and 2249 cyclic pentapeptides that we generated using 14 of the 20 naturally occurring L-amino acids. The cyclic peptides (CPs) dataset contains only head-to-tail cyclic peptides, *i.e.* cyclized from the N-terminus to the C-terminus, thus giving a set of 12-membered rings (see Table S9).

Table S9: Fourteen of the naturally occurring amino acids were used to generate the cyclic peptides dataset.

Property	Amino Acids
Special	Cysteine, Glycine
Charged	Histidine, Lysine, Aspartic Acid, Glutamic Acid
Polar Neutral	Serine, Threonine
Hydrophobic	Alanine, Valine, Leucine, Phenylalanine, Tyrosine, Tryptophan

Figure S2 shows the heavy atoms counts, molecular weights and the rings sizes of the molecules from COD and ZINC. They normally consists of twenty to forty heavy atoms, with few large molecules containing more than 100 heavy atoms. Their molecular weight ranges from 56 dalton to 2039 dalton, with median 386 dalton. Five and six-membered rings are commonly observed ($\sim 10^5$). RDKit[S6] was used to computed these values.

To compare the puckering preference in gas phase and solid state, the full set experimental determined X-ray crystal structures from COD (110197 molecules) were used. These molecules contain a variety of element types in the rings, including boron, carbon, nitrogen, oxygen, silicon, phosphorus, sulphur and selenium.

Distribution of Properties

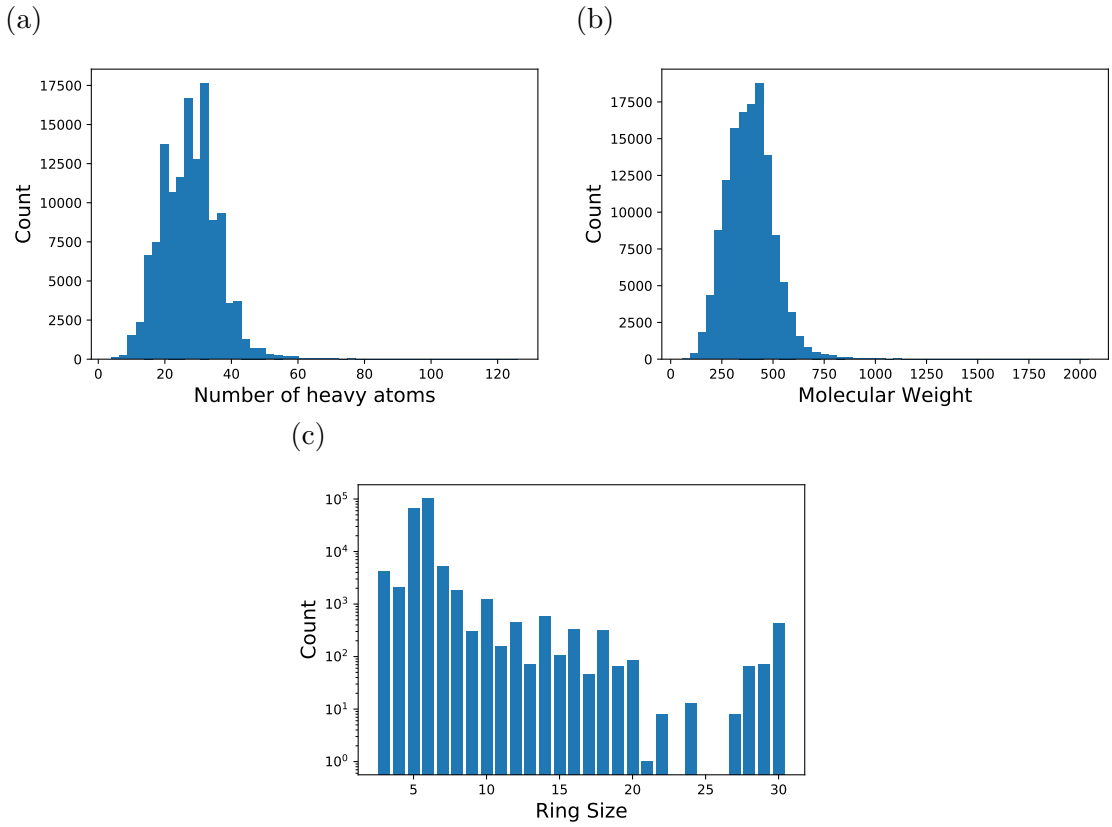


Figure S2: Distribution of Molecular Properties: (a) number of heavy atoms; (b) molecular weight; and (c) distribution of rings by ring size, on a logarithmic scale. Five and six-membered rings are the most common ($> 10^5$) ring sizes.

Table S10: Molecules in the benchmarks set.

Name	SMILES
cyclopentane	C1CCCC1
cyclopentanol	C1CCCC1O
cyclopentene	C1=CCCC1
2-methylcyclopent-2-en-1-ol	CC1=CCCC1O
cyclopent-2-ene-1,5-dione	C1C=CC(=O)C1=O
cyclohexane	C1CCCCC1
methylcyclohexane	CC1CCCCC1
4,4-dimethylcyclohexanone	O=C1CCC(C)(C)CC1
cyclohexene	C1=CCCCC1
1-methyl-1-cyclohexene	CC1=CCCCC1
cyclohexa-1,3-diene	C1=CC=CCC1
cyclohexa-1,4-diene	C1=CCC=CC1
2,5-cyclohexadienone	C1C=CC(=O)C=C1
cycloheptane	C1CCCCCC1
cycloheptanone	O=C1CCCCCC1
cycloheptene	C1=CCCCCC1
cyclohept-2-en-1-one	O=C1C=CCCCC1
cyclohepta-1,3-diene	C1=CC=CCCC1
cyclohepta-1,4-diene	C1=CCC=CCC1
6-methylcyclohepta-1,4-diene	C1=CCC=CCC1C

Generating Low Energy Structures

Experimental-Torsion Distance Geometry with basic Knowledge[S7] implemented in RDKit was use to generate initial conformations, and Open Babel (version 2.4)[S8] was used for the conversion between SDF file to XYZ file.

We simulated the lowest energy conformation using the Conformer-Rotamer Ensemble Sampling Tool[S9] (CREST) based on GFN-xtb energy function.[S10,S11] iMTD-GC workflow was used in the search. Note that CREST may break molecules into fragments, and these molecules are excluded in our analysis.

Ring Conformational Preferences

To investigate the puckering preferences of general rings, we computed all the relevant cycles using RingDecomposerLib,[S12] and the Cremer-Pople puckering parameters were calculated for each relevant cycle. For each ring size N, we clustered the rings according to the number

of non-single bonds in the ring, *i.e.* endocyclic double bonds, shared aromatic bonds and triple bonds. The puckering preferences for different ring sizes are summarized below.

The canonical forms of small (5- and 6-membered) rings are well known, namely the envelope and half-chair conformations for 5-membered rings; and chair, twist-chair, half-chair, and boat conformations for 6-membered rings. These canonical forms give different puckering amplitudes, see Figure S3 and S4. There is one pseudorotation phase angle in both cases, and they are completely free when the rings do not contain any double bonds or shared aromatic bonds.

The presence of endocyclic double and shared aromatic bonds typically alters the conformational preference. To assess the effect of additional constraints on puckering preference, we clustered the observations based on the number of non-single bonds in the ring. Furthermore, we separated the rings with more than one non-single bond according to the location of the non-single bonds. The puckering is restricted by the double bond, as illustrated in Figure S3b and S4b. Strong amplitude-phase coupling is observed in 6-membered rings. The locations of the double bonds affect the puckering preferences, as illustrated in Figure S4c.

This framework can be applied to larger rings. Similar to smaller rings, there are multiple conformational clusters in 7-membered rings, as illustrated in Figure S5. There is now an additional phase angle, ϕ_3, and we can observe strong phase-phase coupling in chair/twist-chair conformations, while the pseudo-rotations are free in other conformational clusters. Figure S6 shows multiple clusters in 7-membered rings with one double bond. The phase couplings vary between clusters. Again, the location of double bonds significantly affects puckering preferences, as illustrated in Figure S7. Figure S8 shows the puckering preferences of 8-membered rings without double bonds. Likewise, there are several clusters and the couplings vary between clusters. The pseudo-rotations are highly restricted in boat-chair conformations, while they are free in crown conformations.

For larger rings, the number of conformational clusters increases, while the coupling between puckering amplitudes and phase angles becomes more complex. It should be noted that small local structural changes may result in significant changes in conformation through transannular repulsion and intramolecular interactions. Although the effect sometimes propagates

via ring strain to distal structural features, it is not completely understood. To gain further insight into long range coupled ring bond rotations, we performed cluster analysis on a set of cyclic peptides.

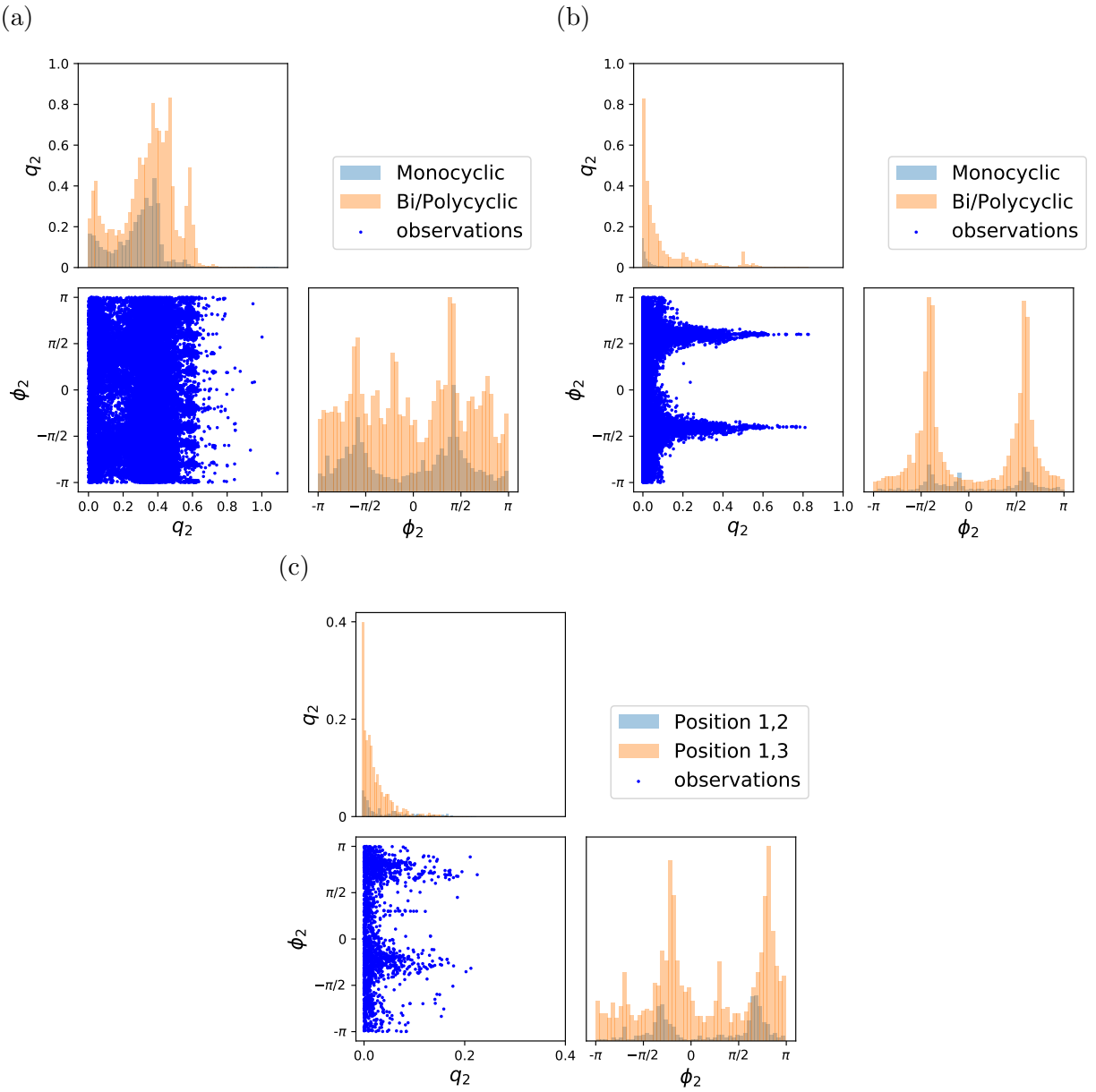


Figure S3: Puckering preferences for 5-membered rings: (a) 5-membered rings without any double bonds or shared aromatic bond; and (b) 5-membered rings with one double bond or shared aromatic bond. Panels (a) and (b) show that the marginal and joint distributions of puckering amplitudes and phase angles in monocyclic rings (light blue), and bi- and polycyclic rings (orange), are similar. Panel (b) shows that the puckering is restricted by the presence of the double bond. (c) 5-membered rings with two double bonds or shared aromatic bonds at 1,2-positions (blue), 1,3 positions (orange). Note that only two aromatic bonds are allowed for positions 1,2. The puckering is further restricted by additional double bonds. The effect of the location of double bonds is small in 5-membered rings.

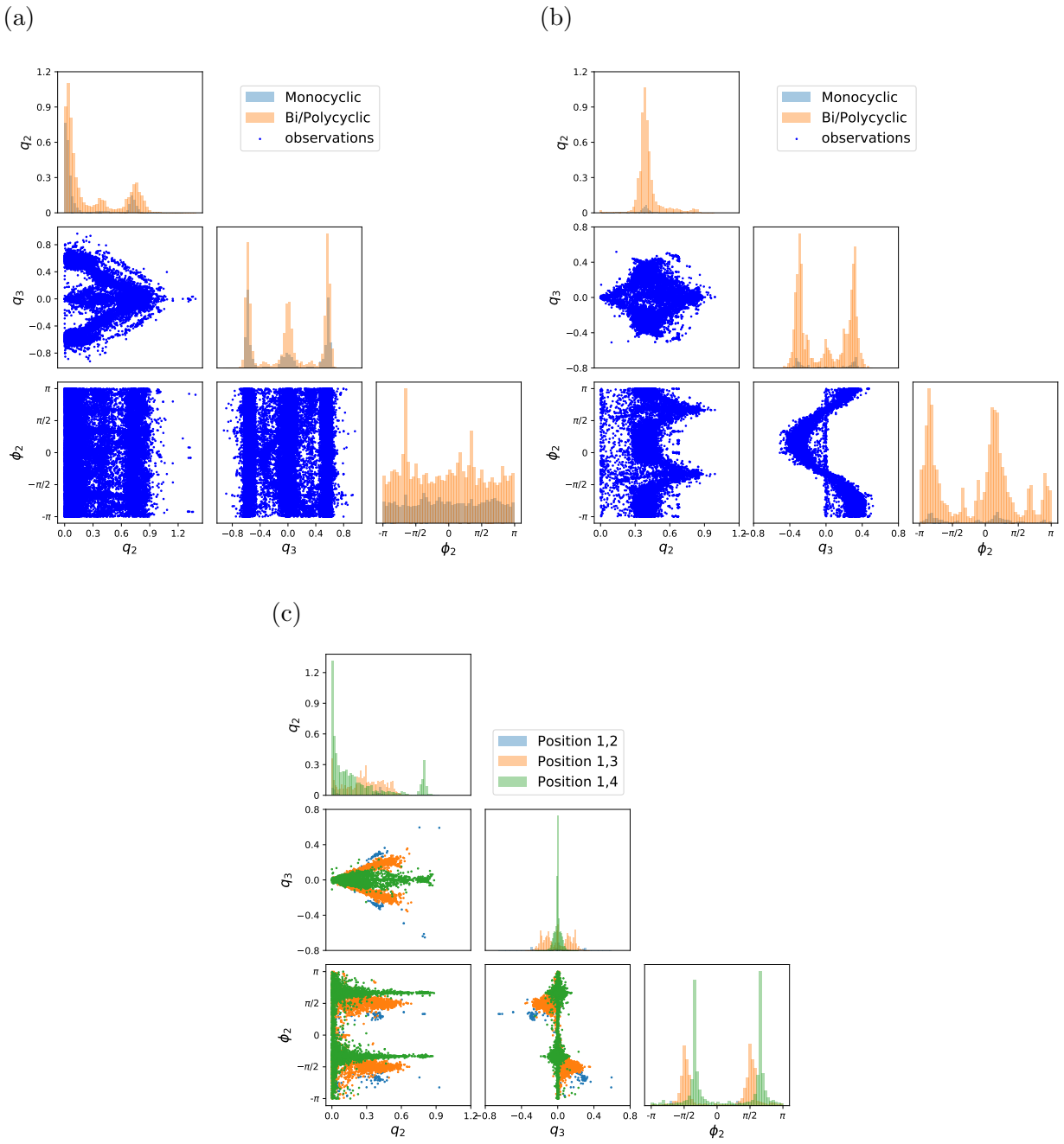


Figure S4: Puckering preferences for 6-membered rings. (a) 6-membered rings without any double or shared aromatic bond. The phase angle, ϕ_2, is uniformly distributed, suggesting that pseudo-rotation is unconstrained. The puckering amplitudes, however, are correlated. The mode ($q_3 = 0$) corresponds to the boat conformation, while the other two modes correspond to the chair conformations. The marginal distribution of puckering amplitudes and phase angles in monocyclic rings (light blue), and bi- and polycyclic rings (orange), are similar. (b) 6-membered rings with one double or shared aromatic bond; the existence of a double bond alters the puckering preferences. Similarly, the marginal distributions of puckering parameters are similar in monocyclic, bi- and polycyclic rings. (c) 6-membered rings with two double or shared aromatic bonds at different positions, as indicated by light blue (positions 1,2; two aromatic bonds only), orange (positions 1,3) and green (positions 1,4). The location of the double or shared aromatic bonds influences the puckering preferences.

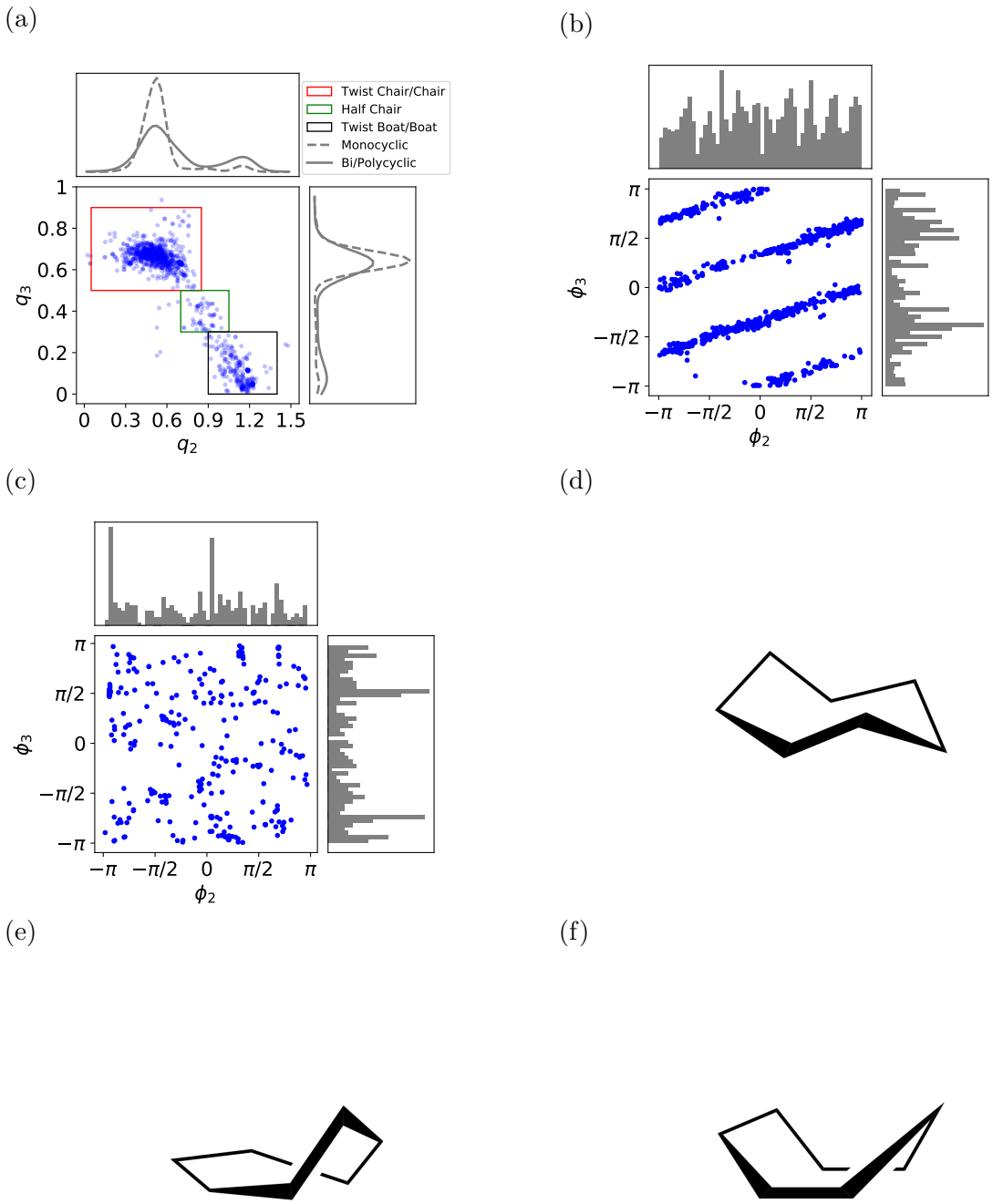


Figure S5: Puckering preferences for 7-membered rings with no double bonds. (a) Puckering amplitudes, q_2 and q_3. The red box indicates the chair/twist-chair region, which has the highest density. There is another mode at $q_2 \approx 1$ and $q_3 \approx 0.1$, which is the boat/twist-boat region (black box). The observations in between are the transitional conformations, *i.e.* half chair, and are less frequently observed (green box). The conformational preferences of both monocyclic rings and complex rings (bi- and poly-cyclic rings) differ slightly, with complex rings having a higher chance of adopting a boat/twist-boat conformation. (b) Phase angles, (ϕ_2, ϕ_3), in the chair/twist-chair region (red box in a). There is strong phase-phase coupling in chair/twist-chair region, while they are marginally uniformly distributed. (c) Phase angles, (ϕ_2, ϕ_3), in the transitional conformation region (green box in a) and boat/twist-boat region (black box in a). There are no clear patterns, which suggests that the pseudorotation is unconstrained in these regions. Conformations of cycloheptane: (d) chair; (e) half chair; and (f) boat.

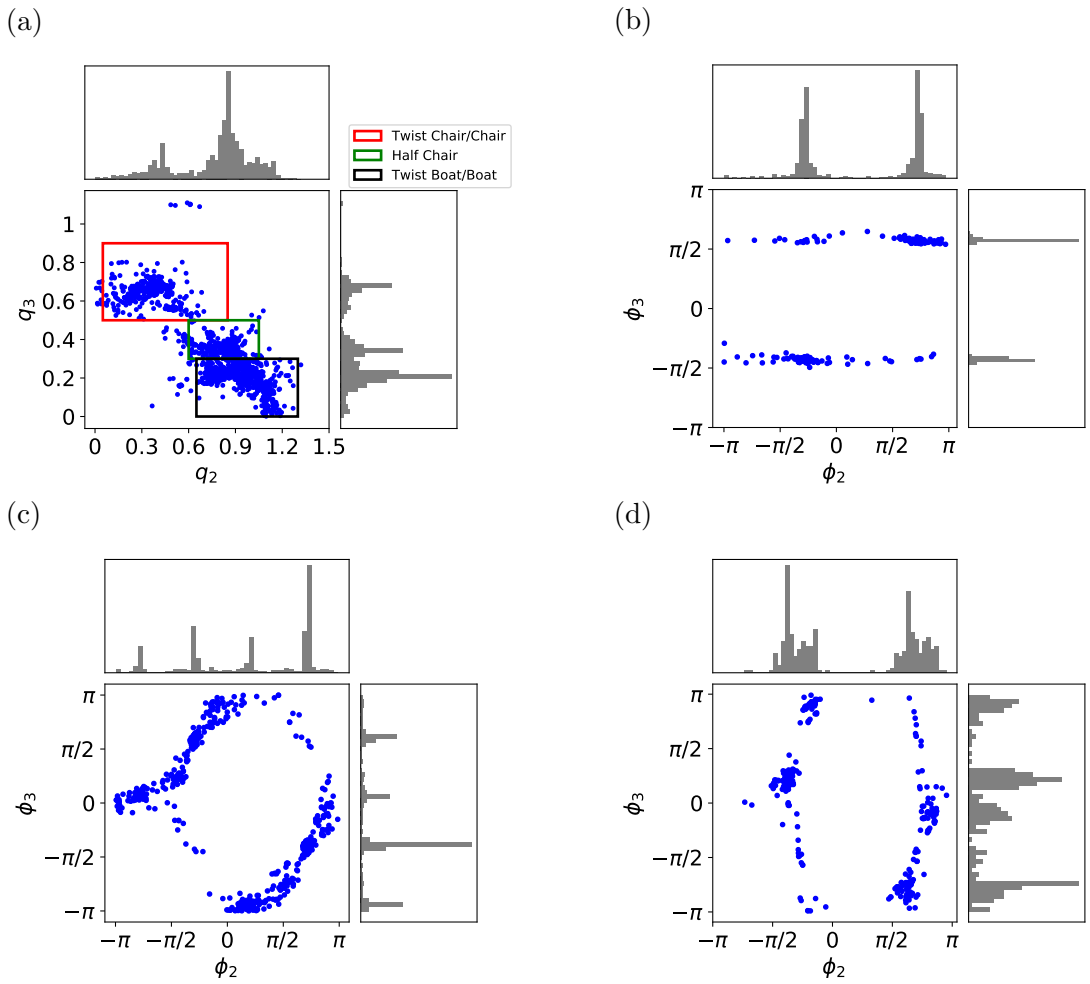


Figure S6: Puckering preferences for 7-membered rings with one endocyclic double or shared aromatic bond. (a) Puckering amplitudes, q_2 and q_3. The red box indicates the chair/twist-chair region. There is another mode at $q_2 \approx 0.85$ and $q_3 \approx 0.2$, which is the boat/twist-boat region. Note that we have combined the conformational preferences of both monocyclic and complex (bi- and poly-cyclic) rings, due to the limited number of observations. Phase angles, (ϕ_2, ϕ_3), in (b) chair/twist-chair region; (c) half-chair conformations and (d) boat/twist-boat region. Figures (b-d) show that pseudo-rotations are generally restricted. In the chair/twist-chair region, ϕ_3 is fixed, while in the boat/twist-boat region, ϕ_2 is fixed. We see strong phase-phase coupling in half chair region.

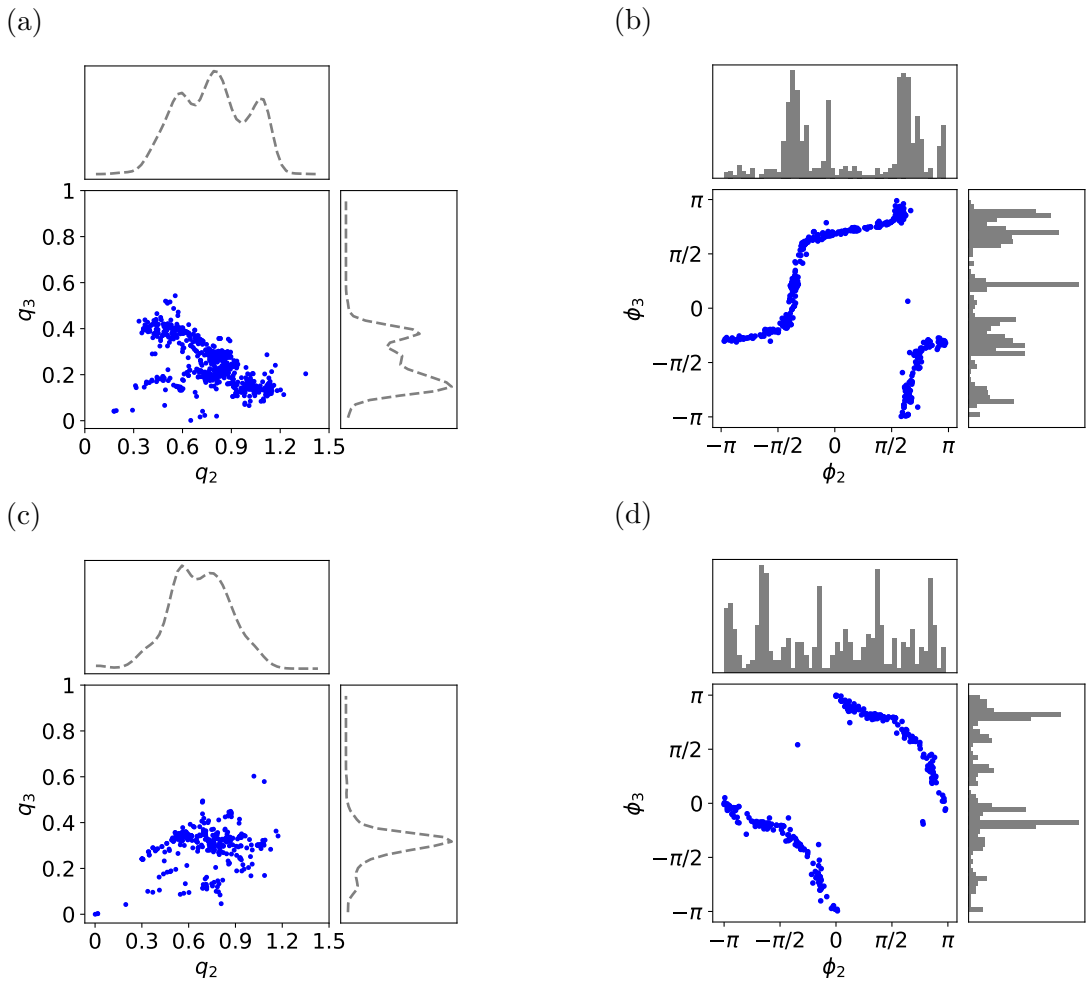


Figure S7: Puckering preferences for 7-membered rings with two endocyclic double or shared aromatic bonds. (a) Puckering amplitudes, q_2 and q_3, in 1,3-cycloheptadiene-like structures. (b) Phase angles (ϕ_2, ϕ_3) in 1,3-cycloheptadiene-like structures. (c) Puckering amplitudes, q_2 and q_3, in 1,3-cycloheptadiene-like structures. (d) Phase angles (ϕ_2, ϕ_3) in 1,4-cycloheptadiene-like structures. Note that we have combined the conformational preferences of both monocyclic and complex (bi- and poly-cyclic) rings, due to the limited number of observations. The location of endocyclic double or shared aromatic bonds has a significant influence on both puckering amplitudes and phase angles. There are different pseudorotation paths in 1,3- and 1,4-cycloheptadiene-like rings.

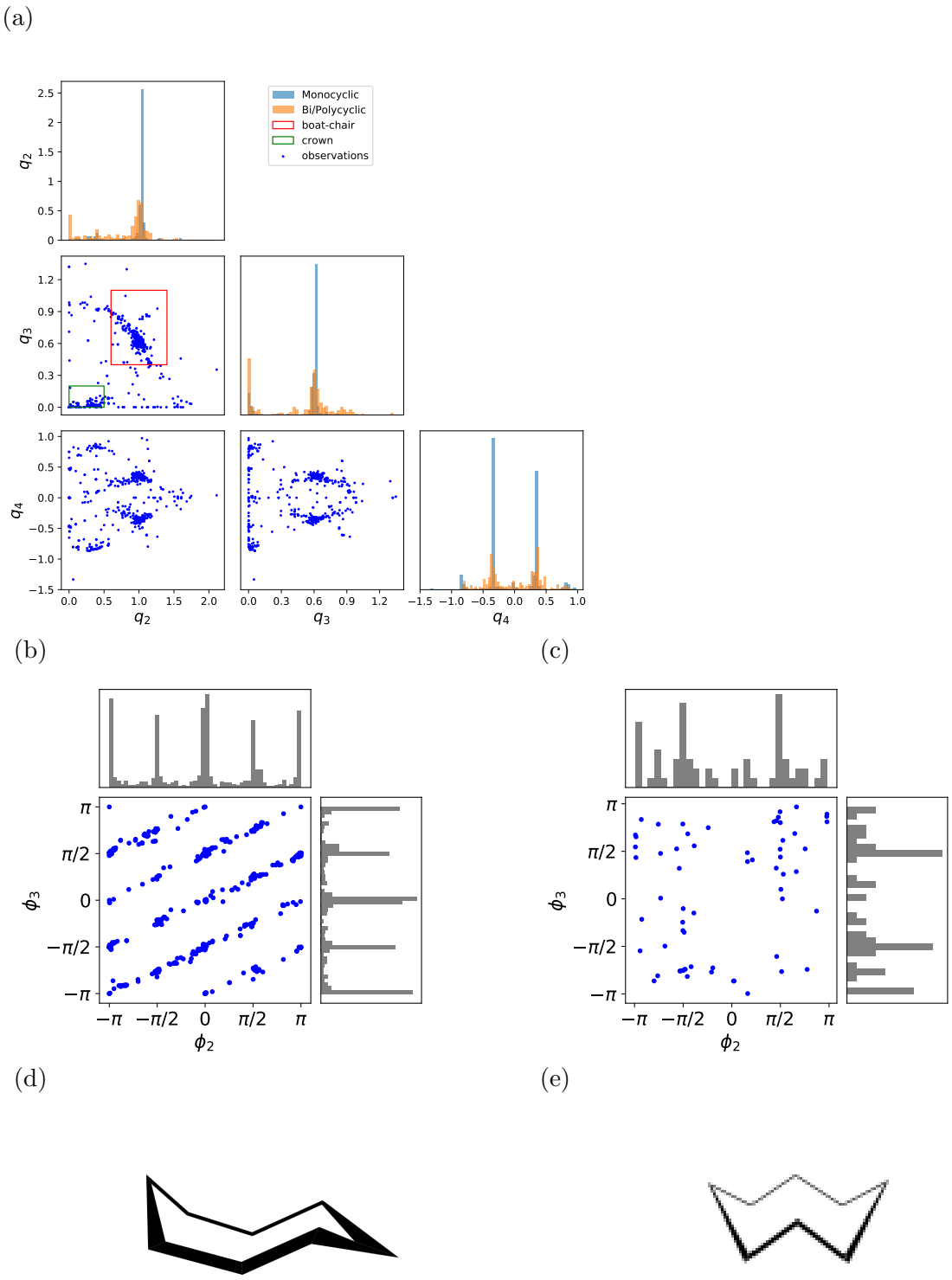


Figure S8: Puckering preferences for 8-membered rings with no double bonds. (a) Puckering amplitudes, (q_2, q_3, q_4); the diagonal histograms show that monocyclic and complex bi- and poly-cyclic rings share similar puckering preferences, and multiple conformational clusters are observed in the scatter plots. The boat-chair conformation (red box) and crown conformation (green box) are frequently observed. (b) Phase angles, (ϕ_2, ϕ_3), for boat-chair conformations. We see strong phase-phase coupling in boat-chair conformations. (c) Phase angles, (ϕ_2, ϕ_3), for crown conformations. There are no clear patterns in panel (c). Conformations of cyclooctane: (d) boat-chair; and (e) crown conformation.

Cyclic Peptides

Peptide cyclization imposes additional constraints on the conformations of peptides, and reduces the thermally-accessible conformational space. There are several factors governing the (backbone) conformation of cyclic peptides, including physicochemical nature of the amino acids (in particular, the steric bulk of side chains), and the formation of intramolecular interactions such as hydrogen bonds. To gain insights into the relative influence of these factors, we investigated the puckering preferences of head-to-tail cyclic tetrapeptides and cyclic pentapeptides.

Similar to small and medium-sized rings, there are multiple canonical conformational clusters in cyclic peptides. These clusters can be classified by the configuration of the amide bonds, thanks to their partial double bond character. The atoms involved in amide bonds are essentially planar, and can adopt either *cis* (C) or *trans* (T) configurations. The configurations of amide bonds leads to distinct types of backbone coupling motions, which are summarized below.

Cyclic Tetrapeptides (Head-to-Tail Cyclization)

There are six possible conformational sequences of amide bonds in cyclic tetrapeptides: (i) TTTT (all *trans*) ; (ii) CTTT ; (iii) CCTT; (iv) CTCT; (v) CCCT; and (vi) CCCC (all *cis*). Typically, the *trans* amide bond is preferred overwhelmingly, however, the high transannular strain reduces the energy barrier between *cis* and *trans* configurations, and *cis* is more favorable in cyclic tetrapeptides. Figure S9a shows that CCCT is the most frequently observed form, followed by CCCC (all-*cis*), and alternating CTCT. These configurations induce distinct geometries, as illustrated in Figure S9b. The multi-modality of eccentricity values indicates multiple sub-clusters within the conformational clusters. They can be classified by the orientation of the amide carbonyl relative to the mean ring plane defined in the Cremer Pople representation *i.e.* either above (U) or below (D) the mean plane. Figure S9c and S9d illustrate the eccentricity values of the rings in each sub-cluster in the all-*cis* (CCCC) and CCCT conformations. In addition to the configuration and orientation of amide bonds, long range intramolecular interactions, such as side chain-side chain interactions, alter the

conformational preferences, and lead to diverse geometries. The orientations of side chain C_β atoms and amide carbonyl groups provide further insights into their effect on intramolecular interactions, see next section.

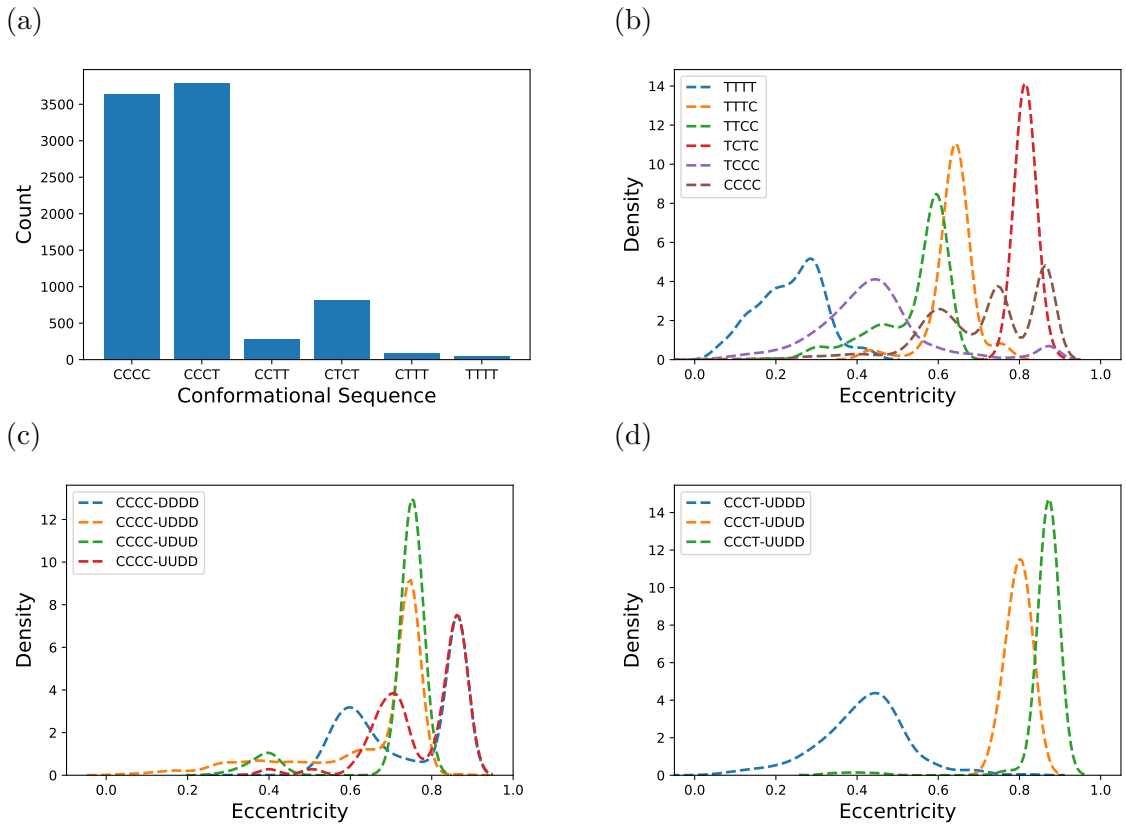


Figure S9: (a) Barplot showing the count of the conformational sequence of amide bonds in the lowest energy conformation of the cyclic tetrapeptides. CCCT is the most frequently observed form, followed by CCCC (all *cis*), and then alternating CTCT. (b) Density plots showing the eccentricity of the backbone under different configurations of the amide bonds. CCCC and alternating CTCT give high eccentricity values, while TTTT (all-*trans*) give the lowest eccentricity values. (c) Eccentricity values of sub-clusters in all-*cis* conformation. (d) Eccentricity values of sub-clusters in CCCT conformation. The orientation of the amide carbonyl is denoted as (U) when it is above the mean plane, while it is denoted as (D) when it is below the mean plane. The orientation amide bonds influence the geometries of the rings. The multimodality in sub-clusters reflects the effect of long range intramolecular interactions on the backbone geometries.

As expected, the puckering preferences vary between different sub-clusters. For example, Figure S10a shows clear-cut puckering amplitudes, (q_2, q_3, q_4, q_5, q_6), in the sub-clusters in all-*cis* conformations. The phase-phase couplings between sub-clusters are also dissimilar: see Figures S10b and S10c. Similar conclusions can be drawn for other conformational clusters. Due to high transannular repulsion, the all-*trans* and CTTT configurations are less favored, and require explicit stabilization through main chain-main chain hydrogen

bonds. The formation of intramolecular hydrogen bonds further restricts their puckering motions (see Figure S11), and results in γ-turns, as illustrated by the Ramachandran plots in Figure S12a. We also calculated the backbone (ϕ,ψ) angles for all clusters. Figure S12b shows that the general torsional preferences are similar to those observed in acyclic secondary protein structures.

(a)

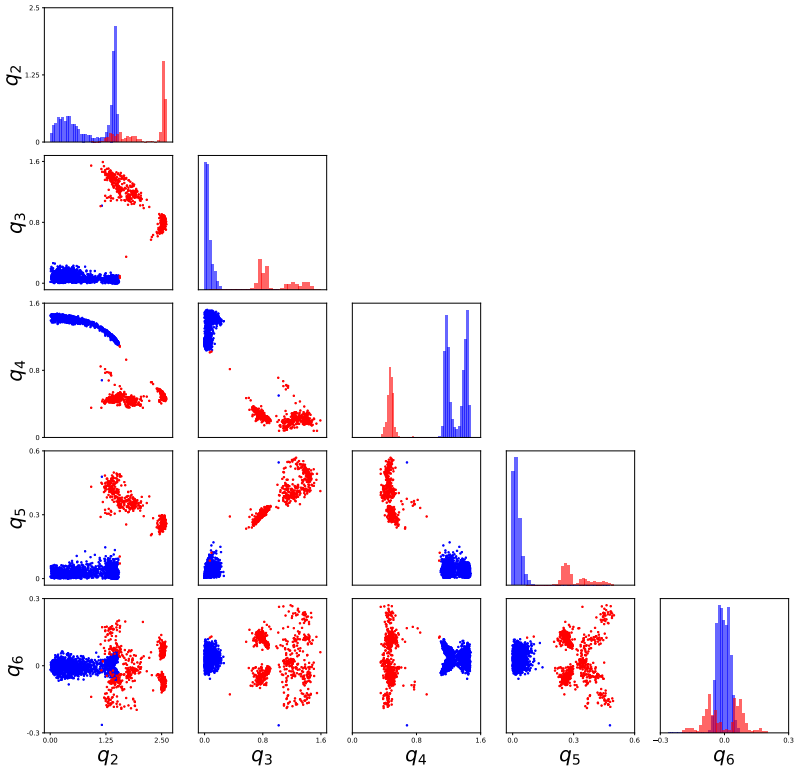
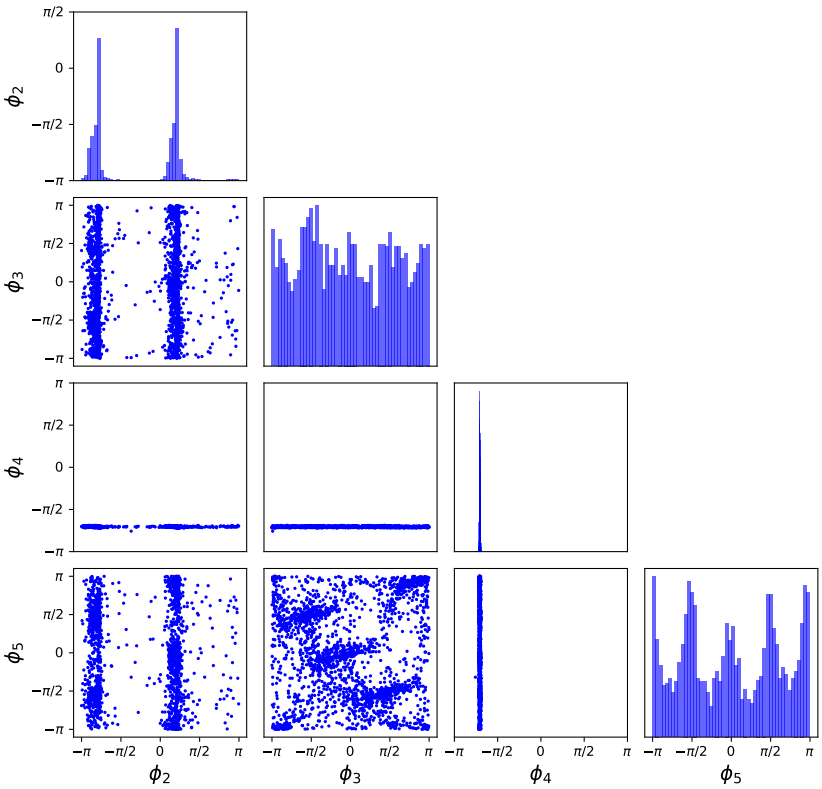


Figure S10: Puckering preferences in all *cis* (CCCC) conformations of cyclic tetrapeptides. (a) Puckering amplitudes, (q_2, q_3, q_4, q_5, q_6), with two sub-clusters, CCCC-DDDD (blue) and CCCC-UDDD (red). Phase angles in (b) CCCC-DDDD, and (c)CCCC-UDDD conformations. The orientation of the amide bonds influence the puckering preferences, and phase-phase couplings vary between sub-clusters.

(b)



(c)

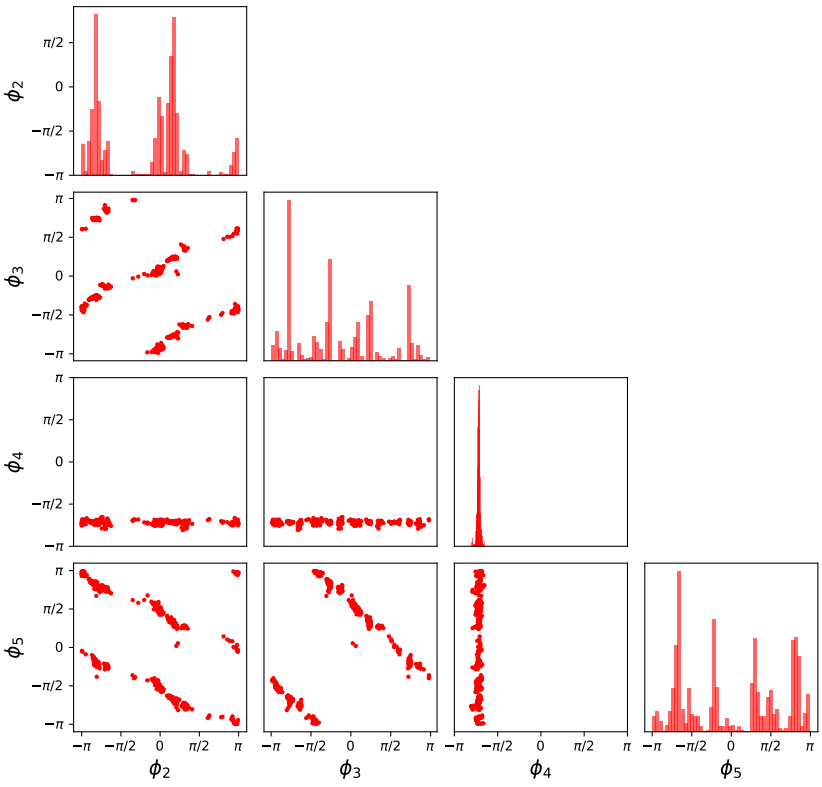
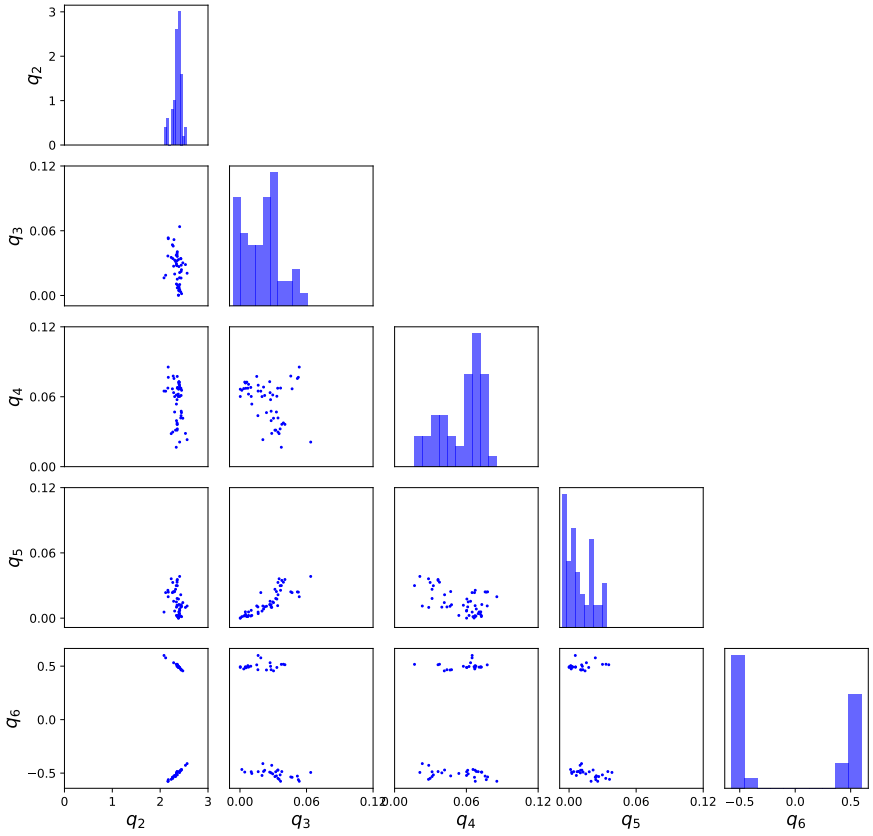


Figure S10: Puckering preferences in all *cis* (CCCC) conformations of cyclic tetrapeptides. (a) Puckering amplitudes, (q_2, q_3, q_4, q_5, q_6), with two sub-clusters, CCCC-DDDD (blue) and CCCC-UDDD (red). Phase angles in (b) CCCC-DDDD, and (c)CCCC-UDDD conformations. The orientation of the amide bonds influence the puckering preferences, and phase-phase couplings vary between sub-clusters.

(a)



(b)

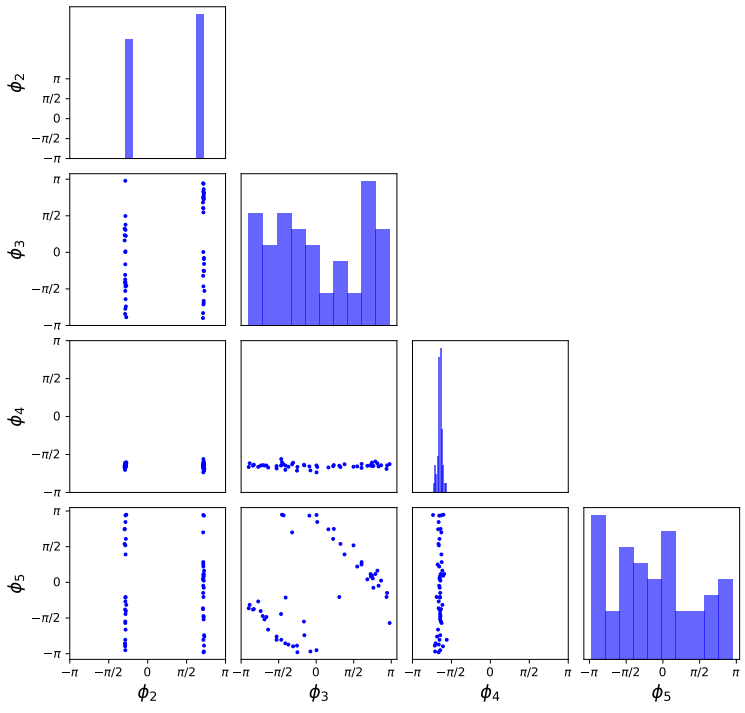


Figure S11: Puckering preferences in all-*trans* (TTTT) conformations of cyclic tetrapeptides. (a) Puckering amplitudes, (q_2, q_3, q_4, q_5, q_6). Phase angles in (b) TTTT-UDUD conformations. The formation of main chain-main chain hydrogen bond(s) further restricts puckering motion.

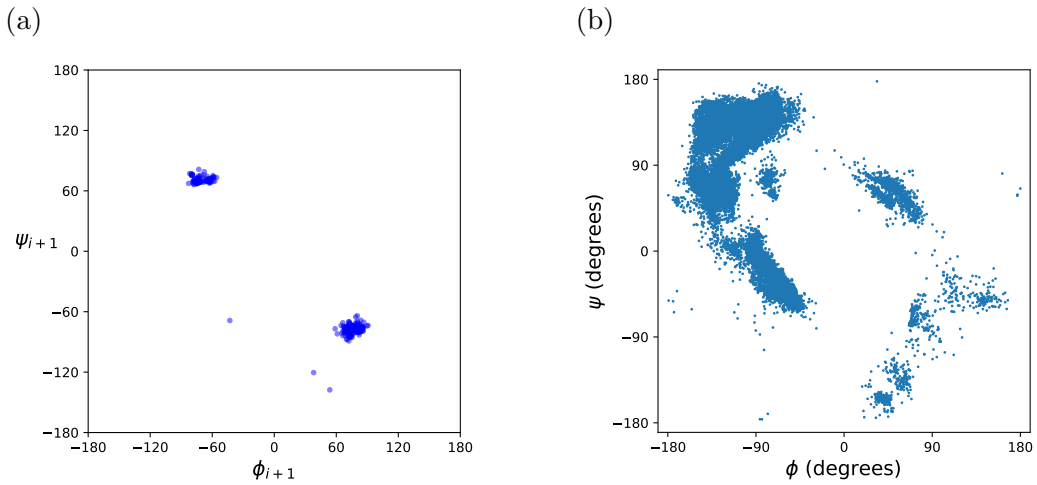


Figure S12: Ramachandran plots showing the backbone (ϕ,ψ) torsion angles at the α-carbon atoms in (a) $i+1$-th amino acid in CTTT and all-*trans*, TTTT, and CTTT conformations in CTPs. The carbonyl C=O in the i-th amino acid and the N-H in the $i+2$-th amino acid form an intramolecular hydrogen bond, and give rise to rigid (ϕ,ψ) torsional preferences at the $i+1$-th amino acid. This results in γ-turns in these two conformational clusters. (b) Backbone (ϕ, ψ) angles for all 14 amino acids (include Gly) in CTPs. The torsional preferences are quite similar to those observed in the secondary structures of proteins.

Cyclic Pentapeptide (Head-to-Tail Cyclization)

We performed a similar analysis for cyclic pentapeptides. Figure S13a shows that the *cis*-amide configuration is more favorable in cyclic pentapeptides. The changes in amide configurations in cyclic pentapeptides lead to diverse geometries, as illustrated in Figure S13b. The orientation of amide carbonyl groups also strongly influences the backbone conformation: see, for example, Figure S13c. Similarly, the multi-modality in sub-clusters reflects the effect of long-range molecular interactions. The puckering amplitudes and phase angles vary between all sub-clusters, as shown in Figure S14. Like cyclic tetrapeptides, the formation of main chain-main chain intramolecular hydrogen bonds in CTTTT and CCTTT cyclic pentapeptide conformations leads to the formation of both β- and γ-turns. The corresponding puckering amplitudes and phase angles are highly restricted.

We calculated the (ϕ,ψ) angles of all amino acids in cyclic pentapeptides. Figure S13d shows the (ϕ,ψ) angles preferences are similar to that in standard protein secondary structures.

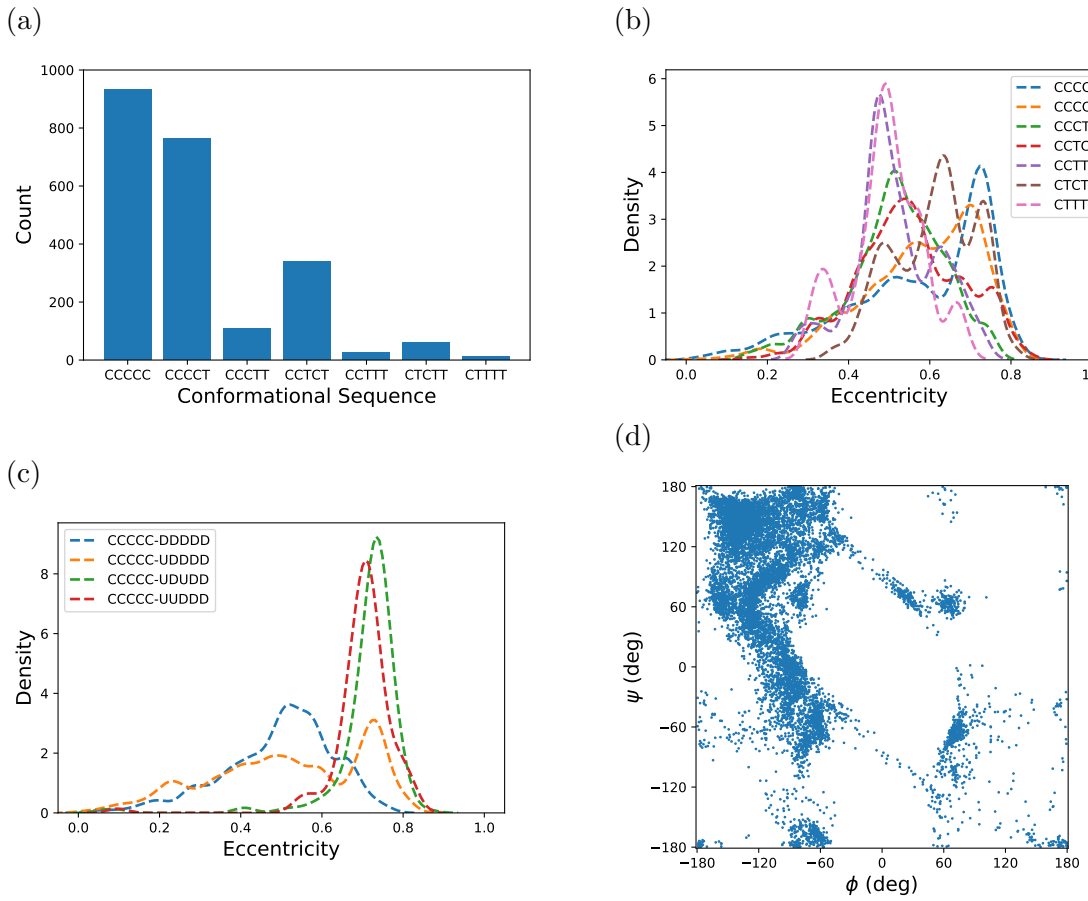


Figure S13: (a) Barplot showing the counts of different types of conformational sequences of amide bonds in the lowest energy conformation of cyclic pentapeptides. All-*cis* amides (CCCCC) is the most frequently observed form, followed by CCCCT and CCTCT. (b) Density plots showing the eccentricity of the cyclic backbone under different configurations of the amide bonds. Different configurations of amide bonds lead to diverse geometries in the backbone conformations. (c) Eccentricity values of sub-clusters in all-*cis* conformation. The orientation of amide bonds influence the backbone geometries. (d) Ramachandran plot for (ϕ,ψ) angles in all amino acids (include Gly) in cyclic pentapeptides. The (ϕ,ψ) angles preferences are similar to that in normal protein secondary structures.

(a)

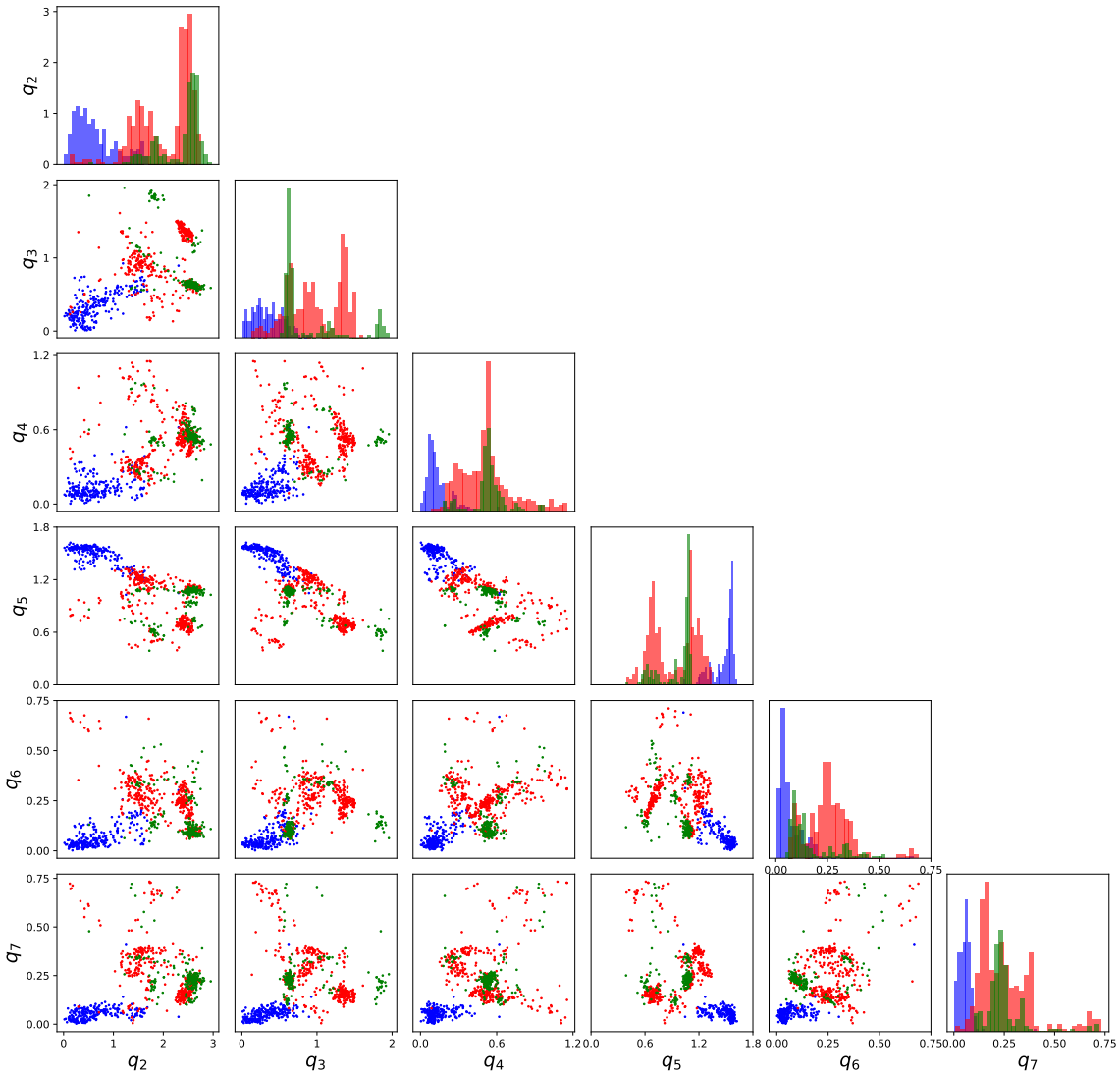


Figure S14: (a) Puckering preferences in all-*cis* conformations of cyclic pentapeptides. (a) Puckering amplitudes, $(q_2, q_3, q_4, q_5, q_6, q_7)$, with three sub-clusters, DDDDD (blue), UDDDD (red), and UDUDD (green) conformaitons. Phase angles in (b) DDDDD; (c) UDDDD; and (d) UDUDD. The orientations of amide carbonyl groups influence the puckering amplitudes and phase angles preferences.

(b)

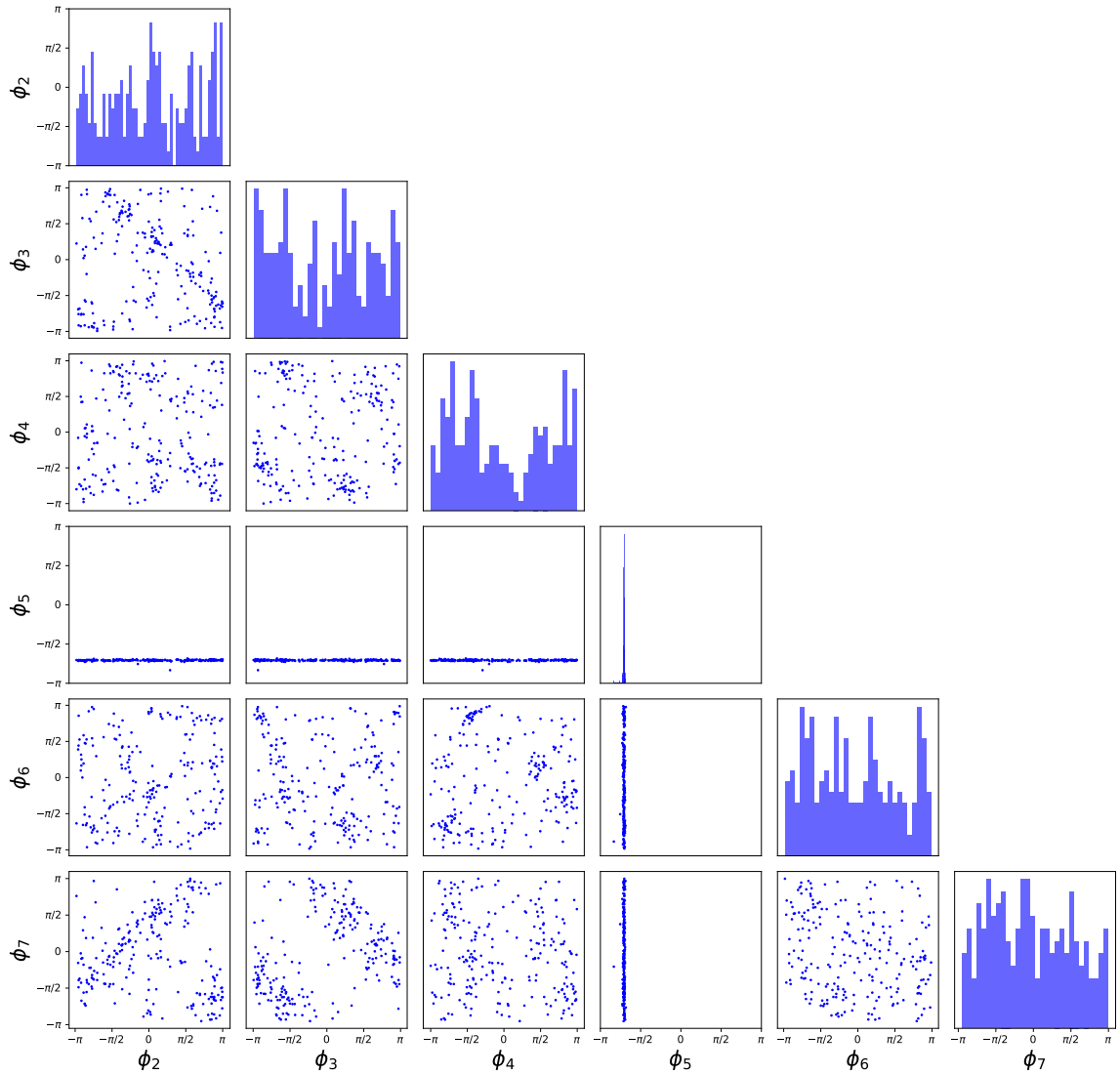


Figure S14: (a) Puckering preferences in all-*cis* conformations of cyclic pentapeptides. (a) Puckering amplitudes, $(q_2, q_3, q_4, q_5, q_6, q_7)$, with three sub-clusters, DDDDD (blue), UDDDD (red), and UDUDD (green) conformaitons. Phase angles in (b) DDDDD; (c) UDDDD; and (d) UDUDD. The orientations of amide carbonyl groups influence the puckering amplitudes and phase angles preferences.

(c)

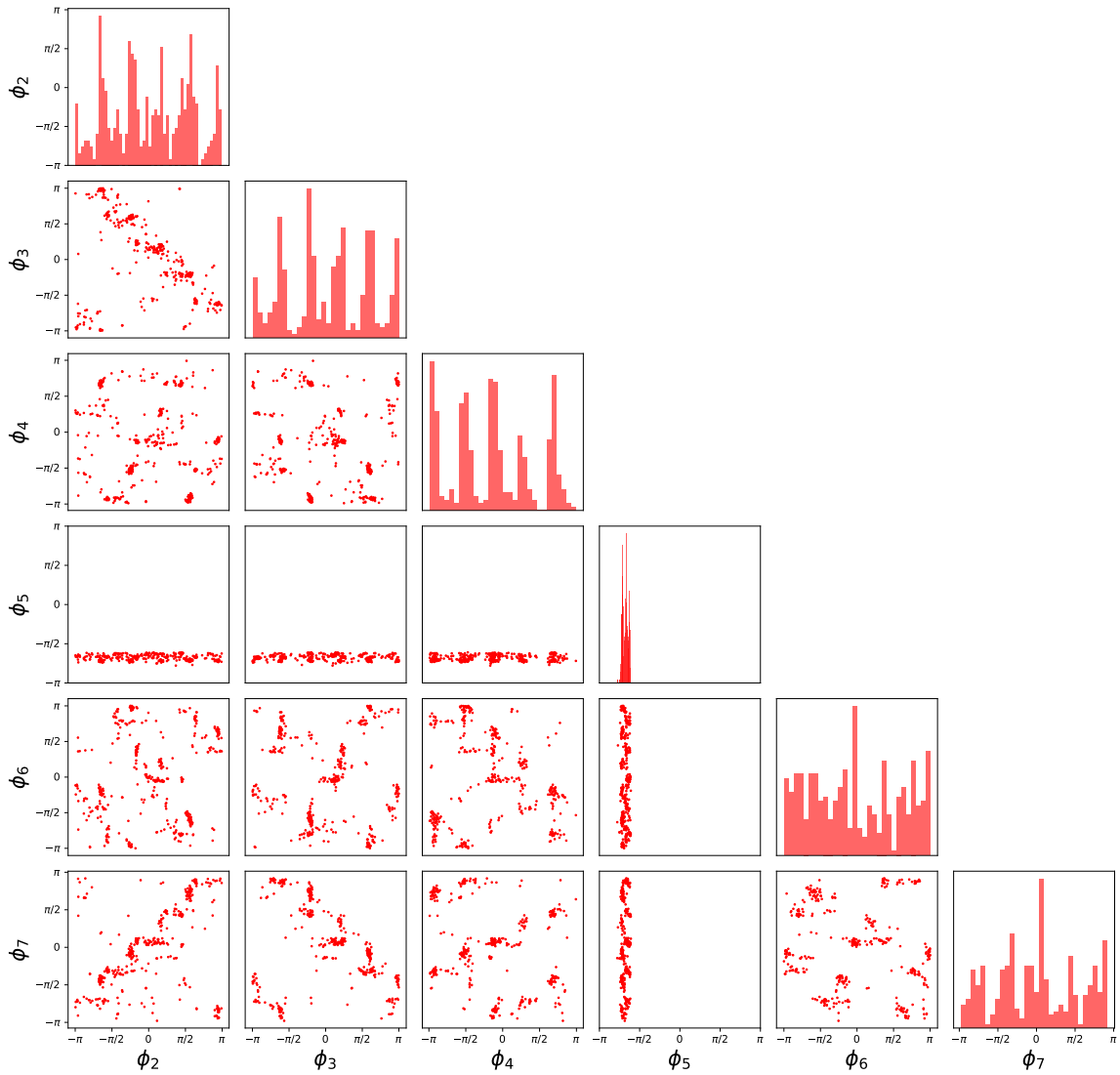


Figure S14: (a) Puckering preferences in all-*cis* conformations of cyclic pentapeptides. (a) Puckering amplitudes, $(q_2, q_3, q_4, q_5, q_6, q_7)$, with three sub-clusters, DDDDD (blue), UDDDD (red), and UDUDD (green) conformaitons. Phase angles in (b) DDDDD; (c) UDDDD; and (d) UDUDD. The orientations of amide carbonyl groups influence the puckering amplitudes and phase angles preferences.

(d)

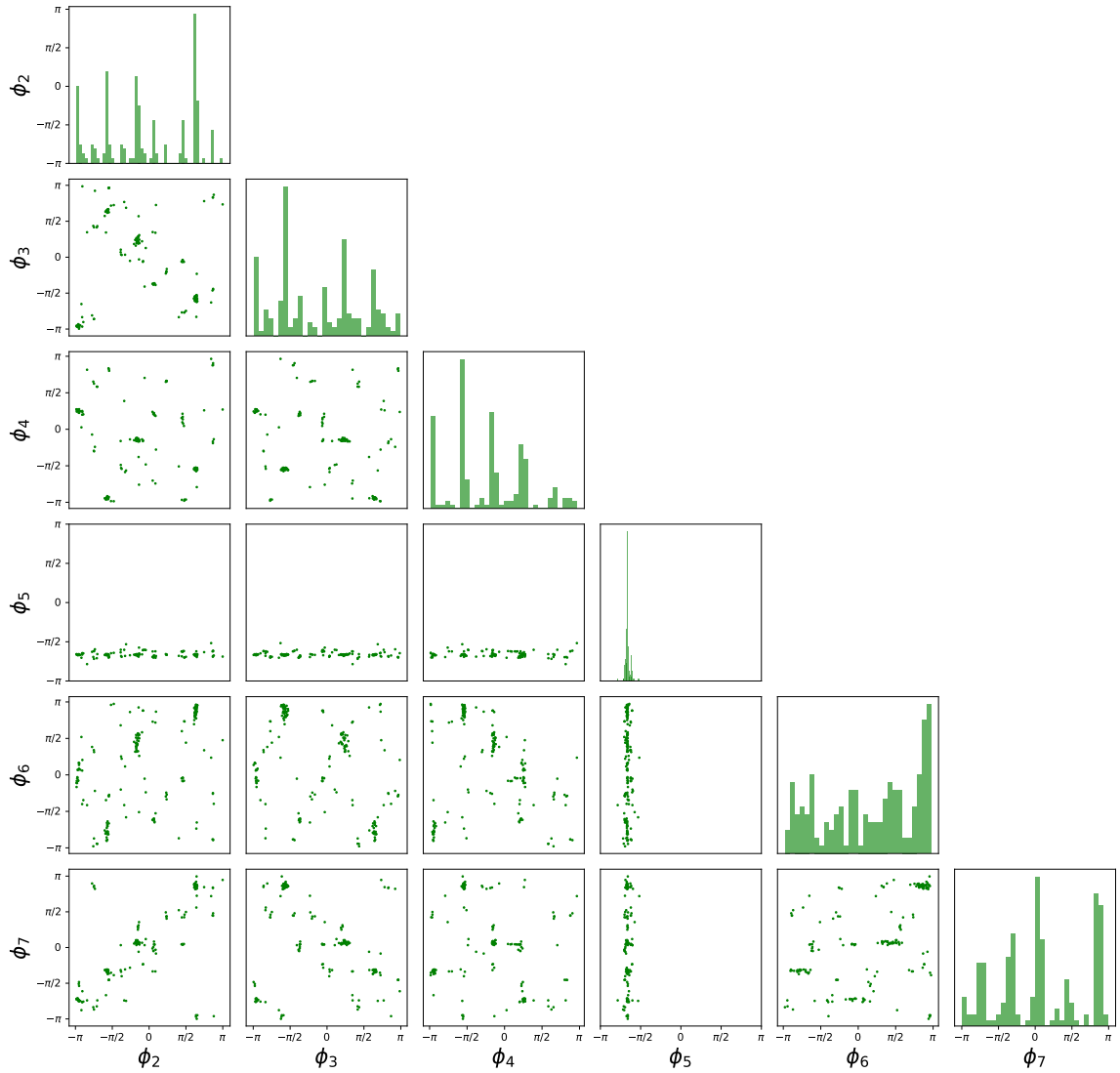


Figure S14: (a) Puckering preferences in all-*cis* conformations of cyclic pentapeptides. (a) Puckering amplitudes, $(q_2, q_3, q_4, q_5, q_6, q_7)$, with three sub-clusters, DDDDD (blue), UDDDD (red), and UDUDD (green) conformaitons. Phase angles in (b) DDDDD; (c) UDDDD; and (d) UDUDD. The orientations of amide carbonyl groups influence the puckering amplitudes and phase angles preferences.

Substituent Position

The size and functionality of substituents are two of the key factors influencing puckering preferences, and their effects vary with ring size. We separated the lowest-energy conformations according to the ring sizes: small rings (5- and 6-membered rings), medium-sized rings (7- to 11-membered rings), and macrocycles (12-membered or larger rings). In particular, we assessed the following substituent types: hydroxyl group, alkoxy group, methyl group, carbonyl group, halogens, and bulky substituents. The bulky substituents is defined by the following SMARTS pattern, [CX4;R]!@;-[CX4H0, CX4H1,CX4H2], where the substituent carbon contains at most 2 hydrogen atoms attached and it is attached to a carbon ring atom with a single non-ring bond.

Figure S15 shows the orientation angle preferences of different substituent types. In small and medium-size rings, the β orientation angle preferences are similar for all substituents types, and they tend to be outwardly directed. However, the α orientation angle changes with ring size. For example, the carbonyl group tends to be equatorial to the mean plane in small rings, and varies in medium-sized rings and macrocycles.

In addition to substituent size, the bond order also influence the substituents orientation. Both endocyclic and exocyclic double bonds restrict the substituent motion, and the substituents tend to be equatorial to the mean plane when it is double bonded or attached to a ring atom linked to a neighbouring ring atom with a shared endocyclic double bond.

In macrocycles, the substituents orientations are widely varying, and small substituents including carbonyl and hydroxyl groups can be inwardly-directed, which are sterically unfavorable in small and medium-sized rings. To gain further insights into substituents orientation preference in macrocycles, we studied the side chain C_β atoms and amide carbonyl positions in cyclic peptides.

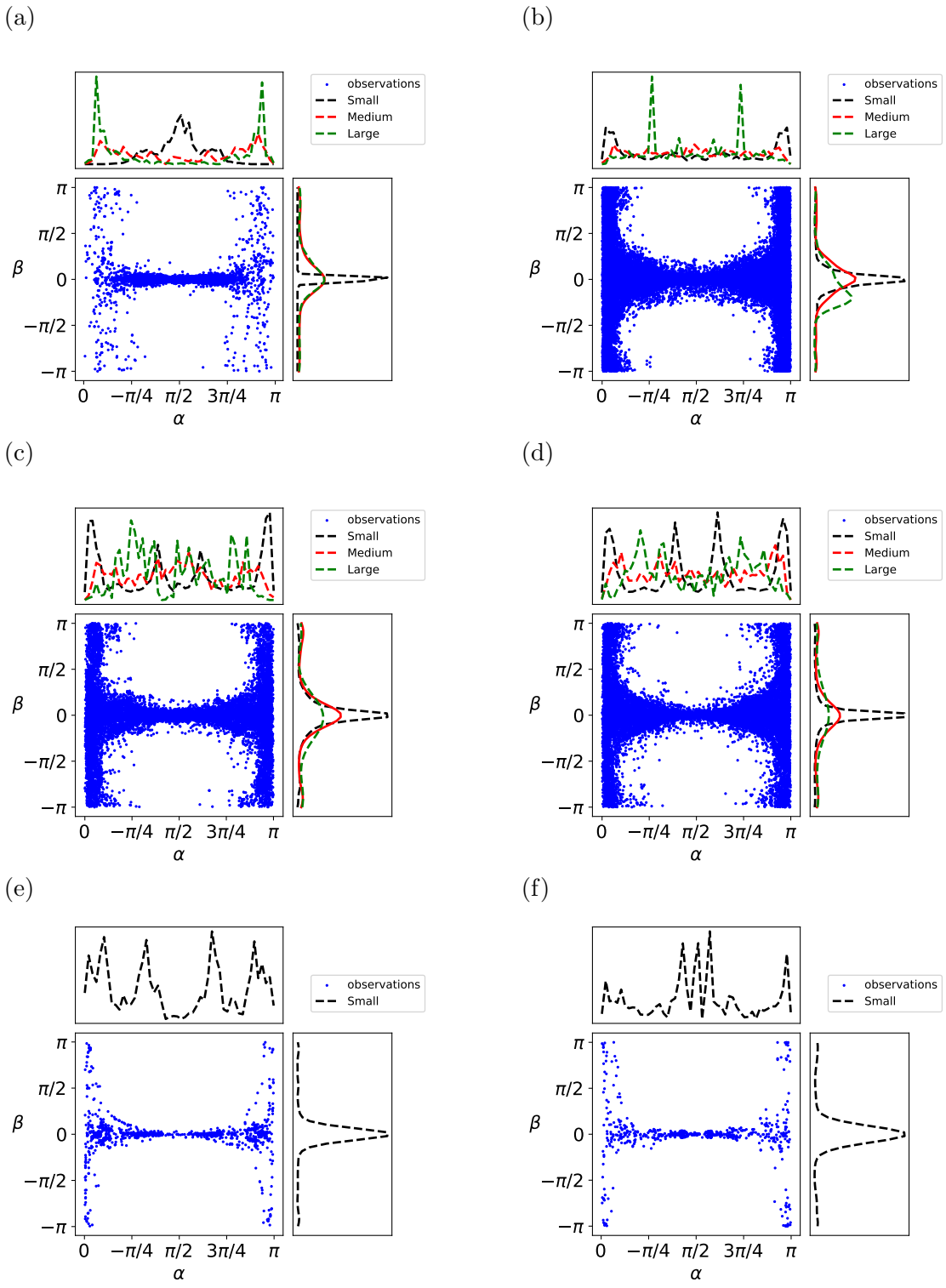


Figure S15: Substituent orientation angle preferences. The marginal distribution of orientation angles of substituents in small, medium-sized, and macrocycles are colored in black, red, and green respectively. (a) carbonyl, (b) methyl (CH_3), (c) alkoxy (d) hydroxyl (e)halogens (fluorine), (f) halogens (chlorine), (g) halogen (bromine), (h) bulky substituents. For halogens, due to small number of observations in medium and large rings, the marginal distribution and the observations are not shown in the figure. All substituents tend to be directed outwardly, as indicated by the β angles. The α orientation angles vary between substituents and ring size.

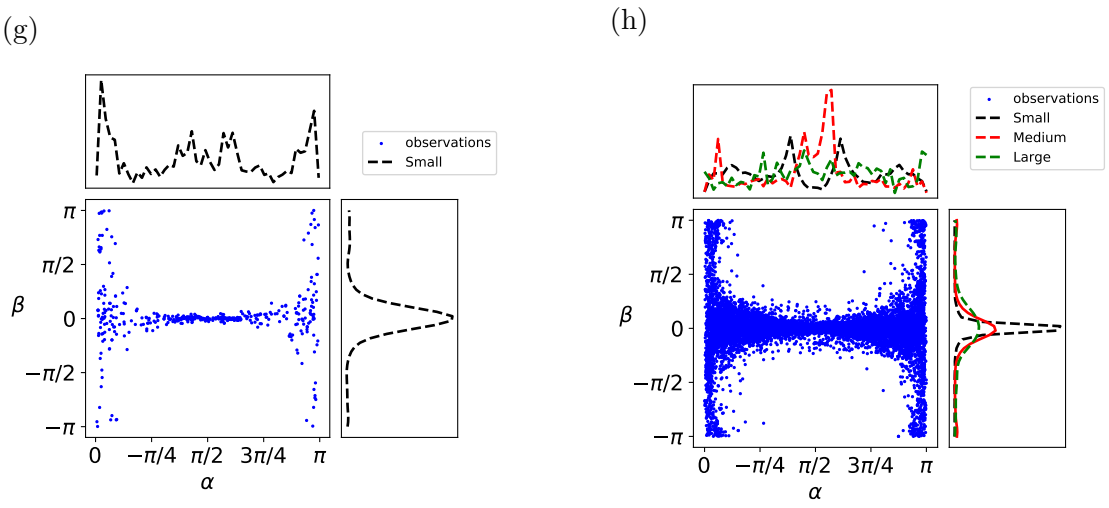


Figure S15: Substituent orientation angle preferences. The marginal distribution of orientation angles of substituents in small, medium-sized, and macrocycles are colored in black, red, and green respectively. (a) carbonyl, (b) methyl (CH_3), (c) alkoxy (d) hydroxyl (e)halogens (fluorine), (f) halogens (chlorine), (g) halogen (bromine), (h) bulky substituents. For halogens, due to small number of observations in medium and large rings, the marginal distribution and the observations are not shown in the figure. All substituents tend to be directed outwardly, as indicated by the β angles. The α orientation angles vary between substituents and ring size.

C_β and Amide Carbonyl Orientation Preference in Cyclic Peptides

As shown in previous section, there are multiple conformational clusters in cyclic peptides, and the puckering preferences vary from clusters to clusters. The long range intramolecular interactions between main chain and side chains influence the puckering preference, and lead to multiple sub-clusters. The substituents orientation angles will help understand the formation of these long range intramolecular interactions.

Figure S16 shows the side chain C_β atoms and amide carbonyl orientation angles preferences in CCCC-DDDD conformational clusters. Both C_β and amide carbonyl groups are generally located accordingly to avoid steric clashes. The sub-clusters orientation angle preferences are colored by blue and green, which explain movement of substituents to align intramolecular interactions. Figure S17 shows the amide carbonyl groups positions are restricted in order to align main chain intramolecular hydrogen bonds in all-*trans* configuration.

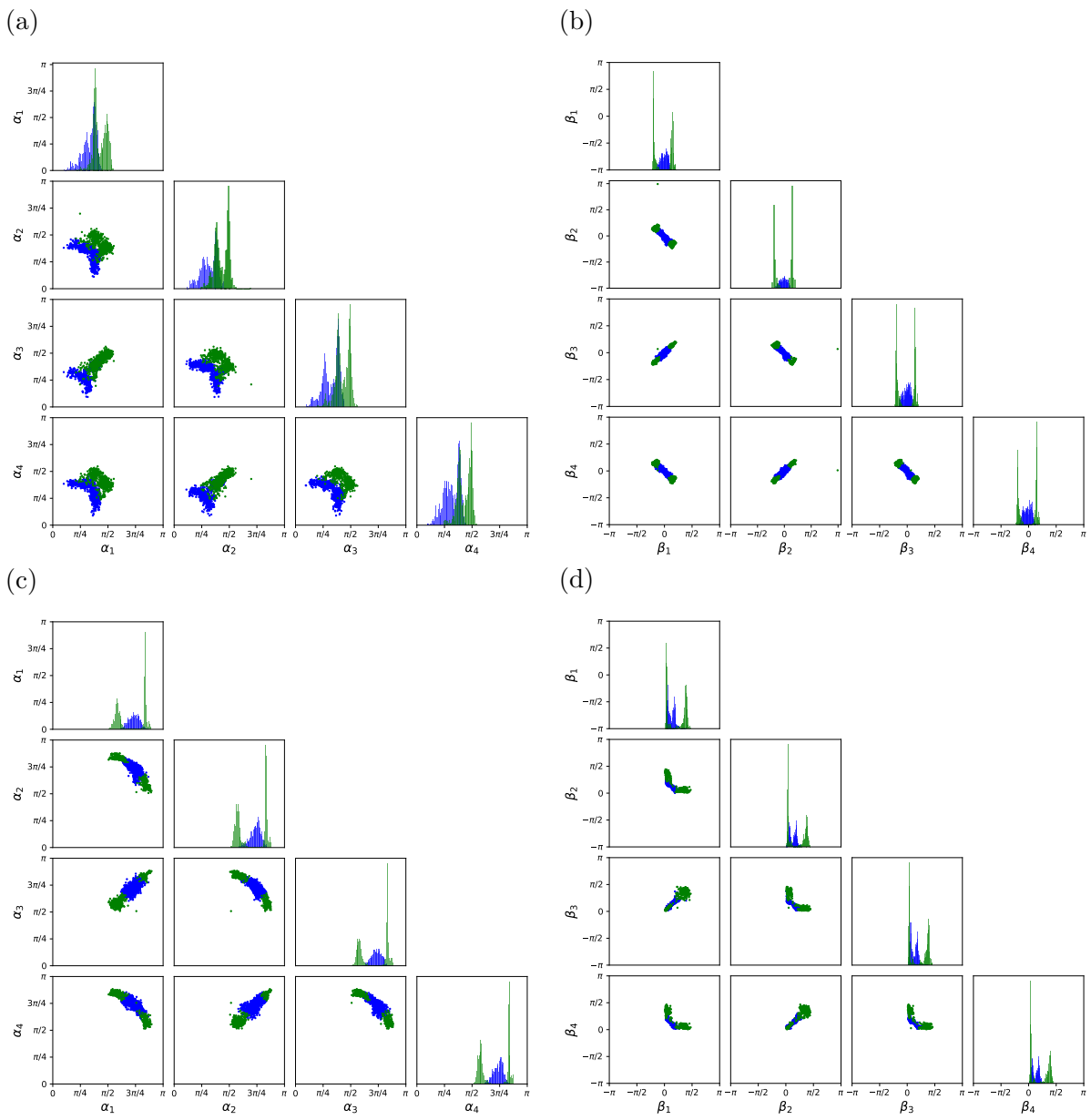


Figure S16: C_β and amide carbonyl orientation preference in cyclic tetrapeptdies with CCCC-DDDD conformations. (a)-(b) α, β angles of C_β. (c)-(d) α, β angles preferences of amide carbonyl. Substituents orientation angles are correlated, and the angles preferences change (colored by blue and green) in order to align the long range intramolecular interactions such as CH-π interactions and hydrogen bonds.

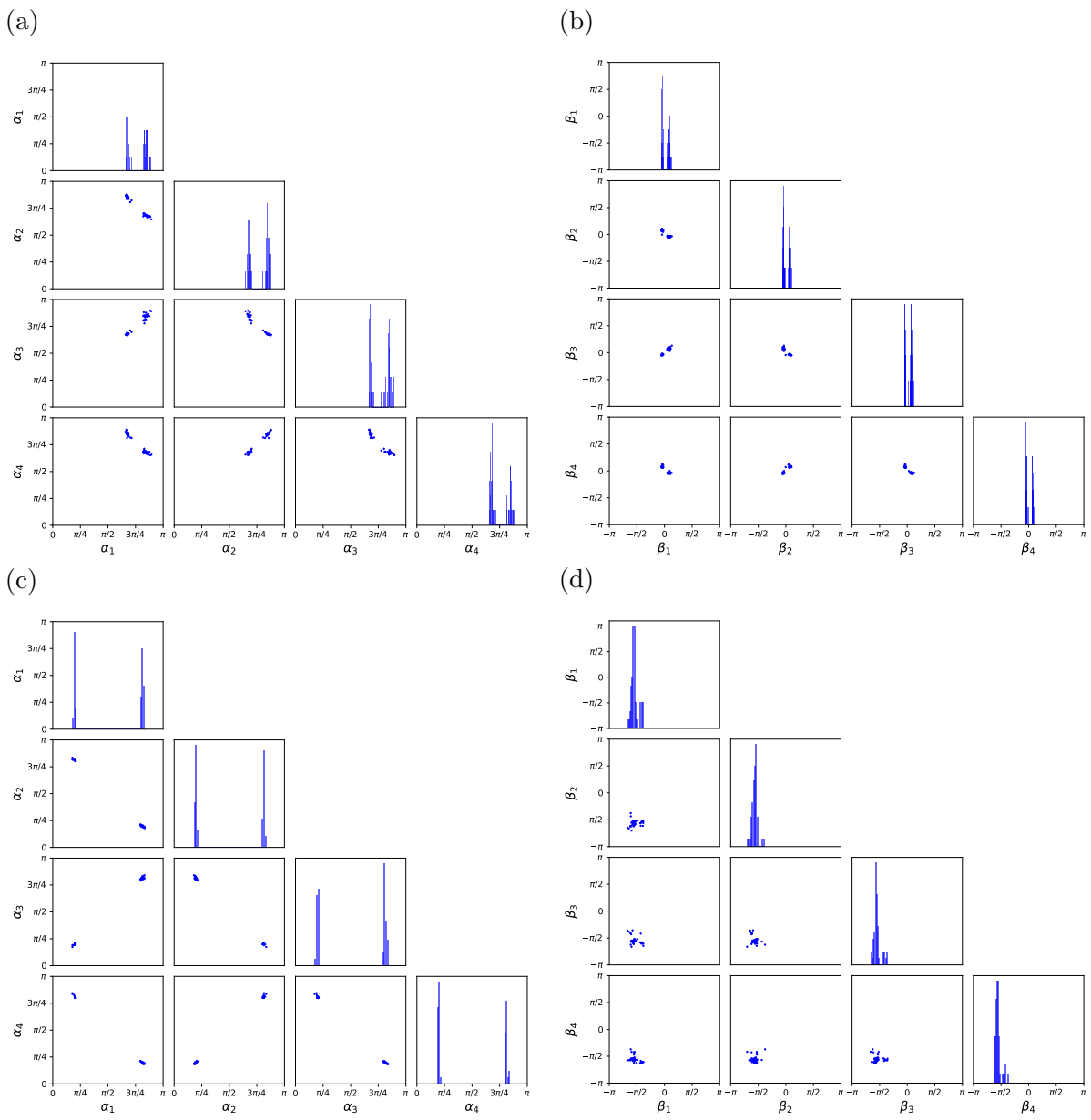


Figure S17: C_β and amide carbonyl orientation preference in cyclic tetrapeptdies with all-*trans* (TTTT) conformations. (a)-(b) α, β angles preferences of C_β. (c)-(d) α, β angles preferences of amide carbonyl. Substituent orientations are restricted in order to align the main chain intramolecular hydrogen bonds.

Side Chain Torsion Angles

In addition to C_β orientation preferences, we also calculated the side chain torsion angles, in particular χ_1 angle. Figure S18 shows multimodal in χ_1 angles for all amino acid side chains, and the torsional preferences are similar to that of protein secondary structures.[S13]

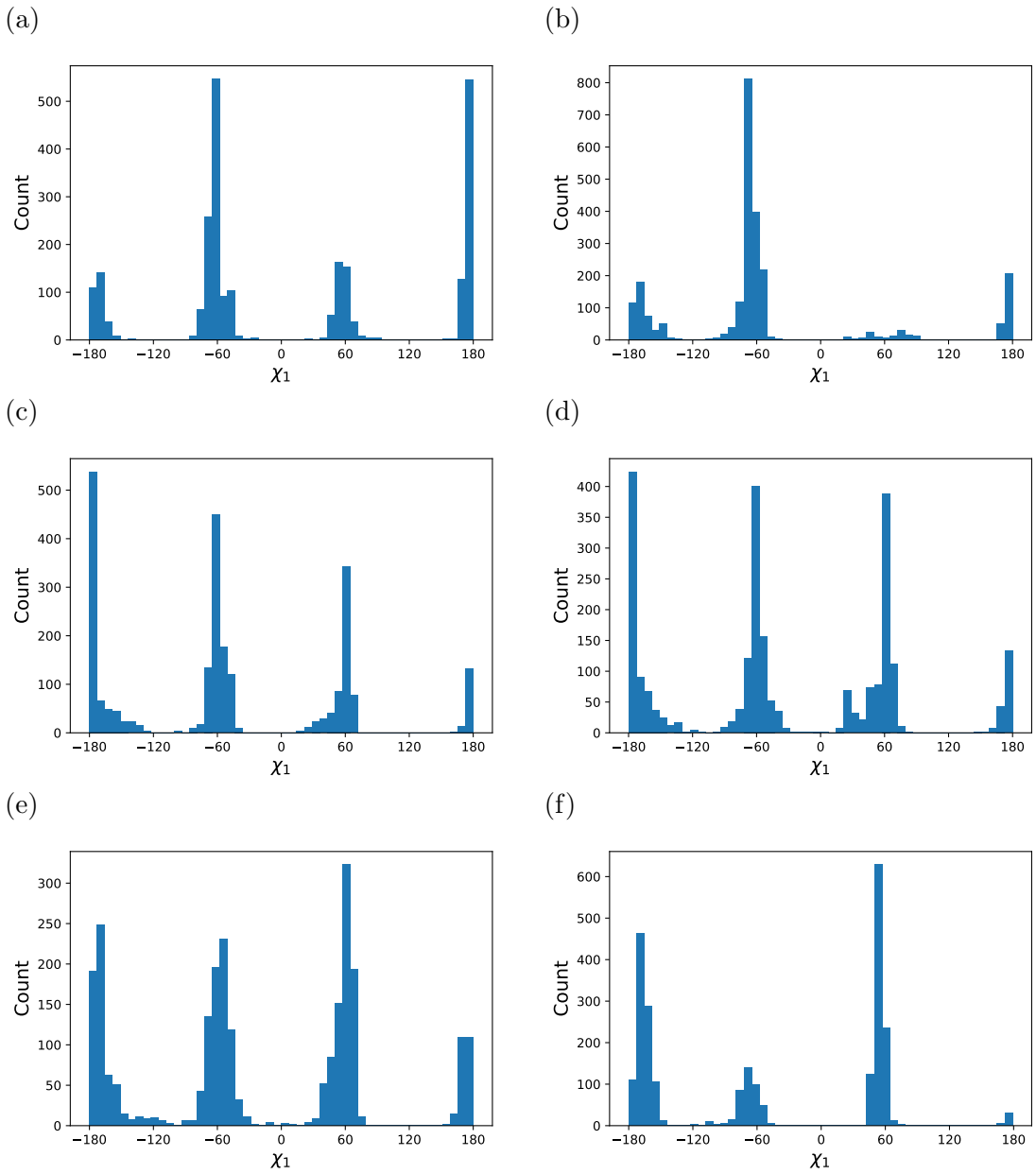


Figure S18: Histogram of χ_1 angles for 12 amino acids. (a) Valine; (b) Leucine; (c) Phenylalanine; (d) Tyrosine; (e) Tryptophan; (f) Cysteine; (g) Serine; (h) Threonine ; (i) Histidine; (j) Lysine; (k) Aspartic acid; (l) Glutamic Acid. The χ_1 angles of all amino acid side chains in CTPs exhibit multimodal, which are similar to the one observed in normal protein side chains.

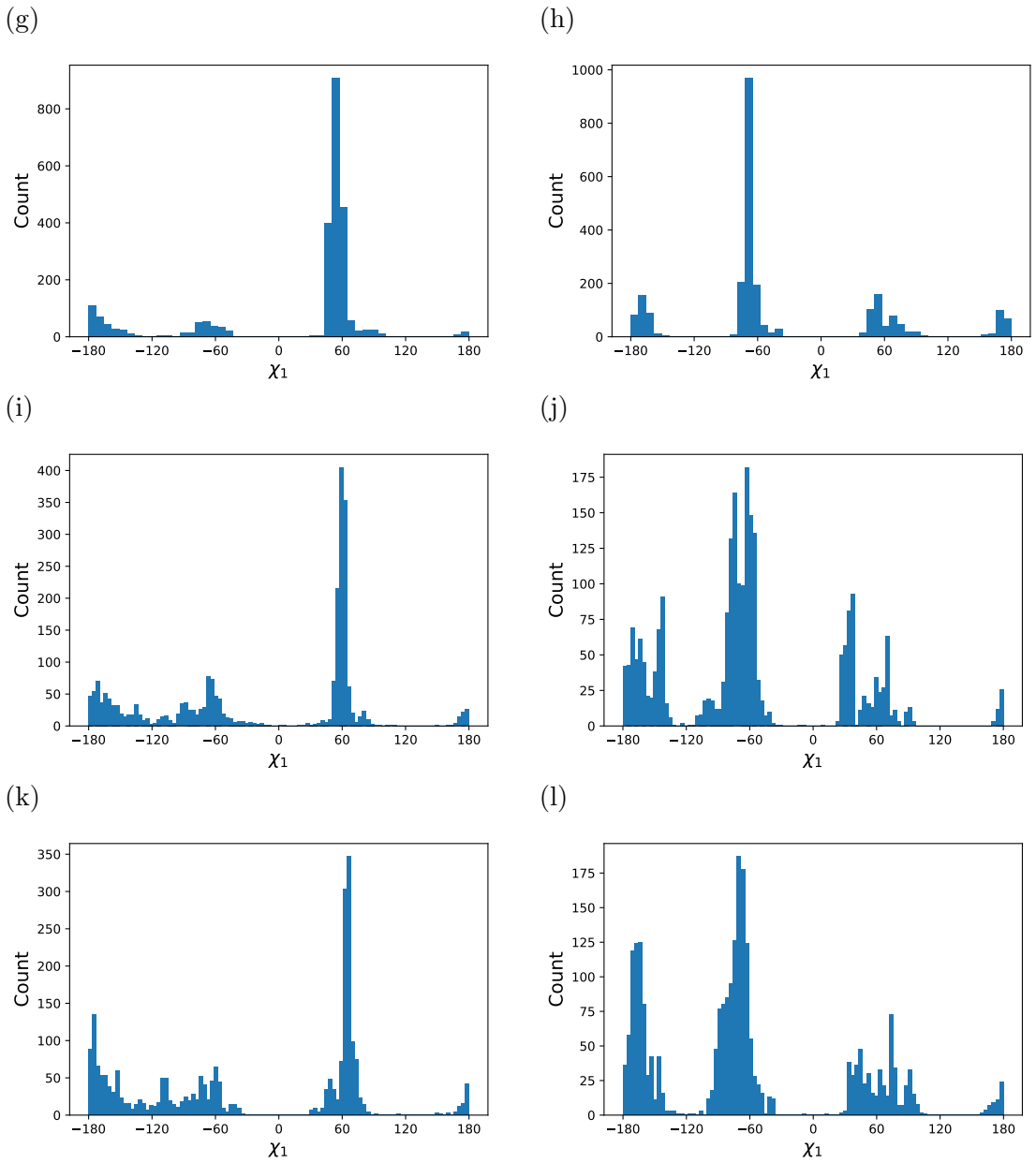


Figure S18: Histogram of χ_1 angles for 12 amino acids. (a) Valine; (b) Leucine; (c) Phenylalanine; (d) Tyrosine; (e) Tryptophan; (f) Cysteine; (g) Serine; (h) Threonine ; (i) Histidine; (j) Lysine; (k) Aspartic acid; (l) Glutamic Acid. The χ_1 angles of all amino acid side chains in CTPs exhibit multimodal, which are similar to the one observed in normal protein side chains.

Connection between Ring Puckering, Substituent Orientation and Torsion Angles

The change in ring substituent positions and ring torsions upon ring puckering are not completely understood. We therefore fitted multiple models to learn the coupling between ring puckering and ring substituent positions, and the torsional changes upon pseudo-rotation. The resulted parameters are listed in Appendix 2, and we provide a summary below.

Ring Puckering and Substituent Orientation

As discussed earlier, the ring conformations are clustered based on its canonical conformations. One could expect the coupling between ring puckering and substituent orientations vary from clusters to clusters. For each cluster, we fitted position-specific models to predict its change in substituent orientation upon pseudo-rotation.

Tables S5 and S6 shows the α and β orientation angles of carbonyl groups in 5-, 6-membered rings and cyclic tetrapeptides in different conformational clusters, at given positions. The predicted α orientation angles is in excellent agreement with the observed angles, with low mean angular error (MAE<0.04) and high squared circular correlation coefficient, $R^2_{circ} > 0.9$. Our models also give low mean angular error (MAE<0.06) in β angle predictions. The low squared circular correlation coefficients are ascribed to the narrow range of β angles in small rings. These models can be applied to large rings with different substituents.

Substituent Exocyclic Torsion Angle

Alternatively, the substituent orientation can be represented by exocyclic torsion angles and exocyclic bond angles. Here we focused on the prediction of exocyclic torsion angles. Since the exocyclic and endocyclic torsion angles are the rotations about same pair of ring atoms $(i,\, i+1)$, Equation (S28) can be directly applied, *i.e.* setting $Q_i = 1$ in Equation (S20), to describe the rotational relationship between substituent position and ring torsions.

$$\theta_i^{exo} = P_i + \theta_i^{endo} \tag{S28}$$

where P_i, θ_i^{exo} and θ_i^{endo} are defined as above.

The parameter P_i depends on the nature of the substituents, and the relative stereochemistry of the stereo-center, and thus require multiple sub-models to describe substituent exocyclic torsion angles. Table S8 shows the prediction of exocyclic torsion angles of different substituents, including alkoxy, hydroxyl, methyl group, halogens, bulky substituents, and carbonyl groups. All sub-models are in excellent agreements with the observed exocyclic torsion angles, with small mean angular error (< 0.06 rad, or equivalently $3.4°$) and high squared circular correlations coefficient, $R^2_{\text{circ}} > 0.97$.

As mentioned earlier, torsion angle is an alternative representation commonly used in the community to study ring geometries. de Leeuw et al. discuss the connection between ring puckering coordinates and torsion angles in small rings, which thus help understand the internal torsional changes upon pseudo-rotations. Here, we proposed general models convert from puckering parameters to endocyclic torsion angles for N-membered rings. Table S7 shows the predictions of endocyclic torsion angles from puckering parameters. The predicted values are in good agreement with the observed torsion angles for both small and large rings, with high square circular correlation coefficient and low mean angular error.

Ring Reconstruction

To assess our proposed method in sampling ring conformers, twenty molecules were selected, as listed in Table S10. The GFN2-computed lowest energy conformation was used as reference conformation. The best RMSD and TFD values of each molecule are reported in Table S11. Note that the conformation with lowest RMSD may not give the lowest TFD values. The TFD values here are computed based on the conformations with the lowest RMSD values. The average RMSD values and the corresponding TFD values are 0.092 Å and 0.048 respectivly. Small TFD values demonstrate the effectiveness of the algorithm. The high RMSD values are ascribed to the deviation in bond length and bond angles, which can be reduced by local geometry optimisation.

Table S11: RMSD and TFD values between sampled conformations and reference conformations. The TFD values are calculated using conformation with the lowest RMSD value.

Molecule	RMSD	TFD
cyclopentane	0.020	0.058
cyclopentanol	0.075	0.167
cyclopentene	0.053	0.039
2-methylcyclopent-2-en-1-ol	0.069	0.013
cyclopent-2-ene-1,5-dione	0.072	0.031
cyclohexane	0.057	0.063
methylcyclohexane	0.111	0.029
4,4-dimethylcyclohexanone	0.164	0.049
cyclohexene	0.073	0.012
1-methyl-1-cyclohexene	0.066	0.067
cyclohexa-1,3-diene	0.079	0.048
cyclohexa-1,4-diene	0.088	0.015
2,5-cyclohexadienone	0.085	0.013
cycloheptane	0.115	0.055
cycloheptanone	0.130	0.007
cycloheptene	0.053	0.003
cyclohept-2-en-1-one	0.166	0.099
cyclohepta-1,3-diene	0.090	0.090
cyclohepta-1,4-diene	0.104	0.044
6-methylcyclohepta-1,4-diene	0.168	0.050

Puckering preference in solid state and gas phase

To gain insights into the puckering preference in solid state, we compared our results with previous empirical studies on crystal structures from Cambridge Structural Database,[S14–S16] and the 63814 experimentally determined X-ray crystal structures from COD. These empirical studies focused on medium sized rings, in particular seven- and eight-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. Figure S19 shows the puckering preference from GFN2-computed low energy conformations (blue) and COD crystal structures (red), in 6-membered rings with one double bond and 7-membered rings with no endocyclic double bond. The marginal and joint puckering preference are similar in both states. Similar observations were made for 5- to 8-membered rings. More data is required to make a fair comparison for larger rings.

Similarly, the substituent orientation angle preferences from GFN2-computed low energy

conformation and the crystal structures from COD are alike in both states, for all substituent types studied above. The proposed models can be directly applied to solid states.

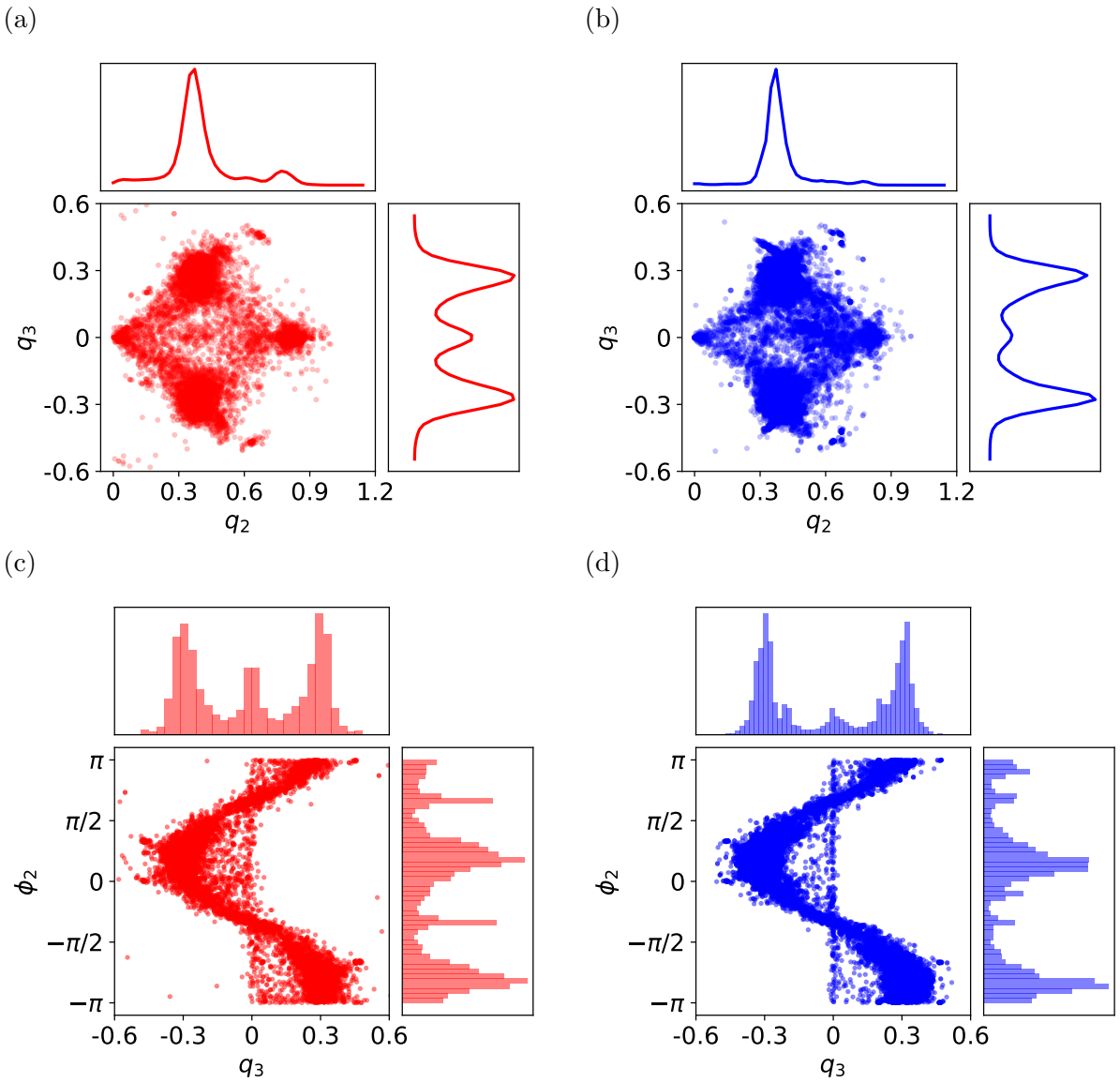


Figure S19: Puckering preferences of 63814 crystal structures from COD (red) and all GFN2-computed low energy conformations (blue). Puckering amplitudes (q_2,q_3) in 6-membered ring with one double bond in (a) solid state; (b) gas phase. Coupled puckering amplitude (q_3) and phase angle (ϕ_2) in 6-membered ring with one double bond in (c) solid state; (d) gas phase. Puckering amplitudes (q_2,q_3) in 7-membered ring with no double bond in (e) solid state; (f) gas phase. Coupled phase angles (ϕ_2, ϕ_3) in chair and twist-chair region (g) solid state; (h) gas phase. The preferences are similar in both states.

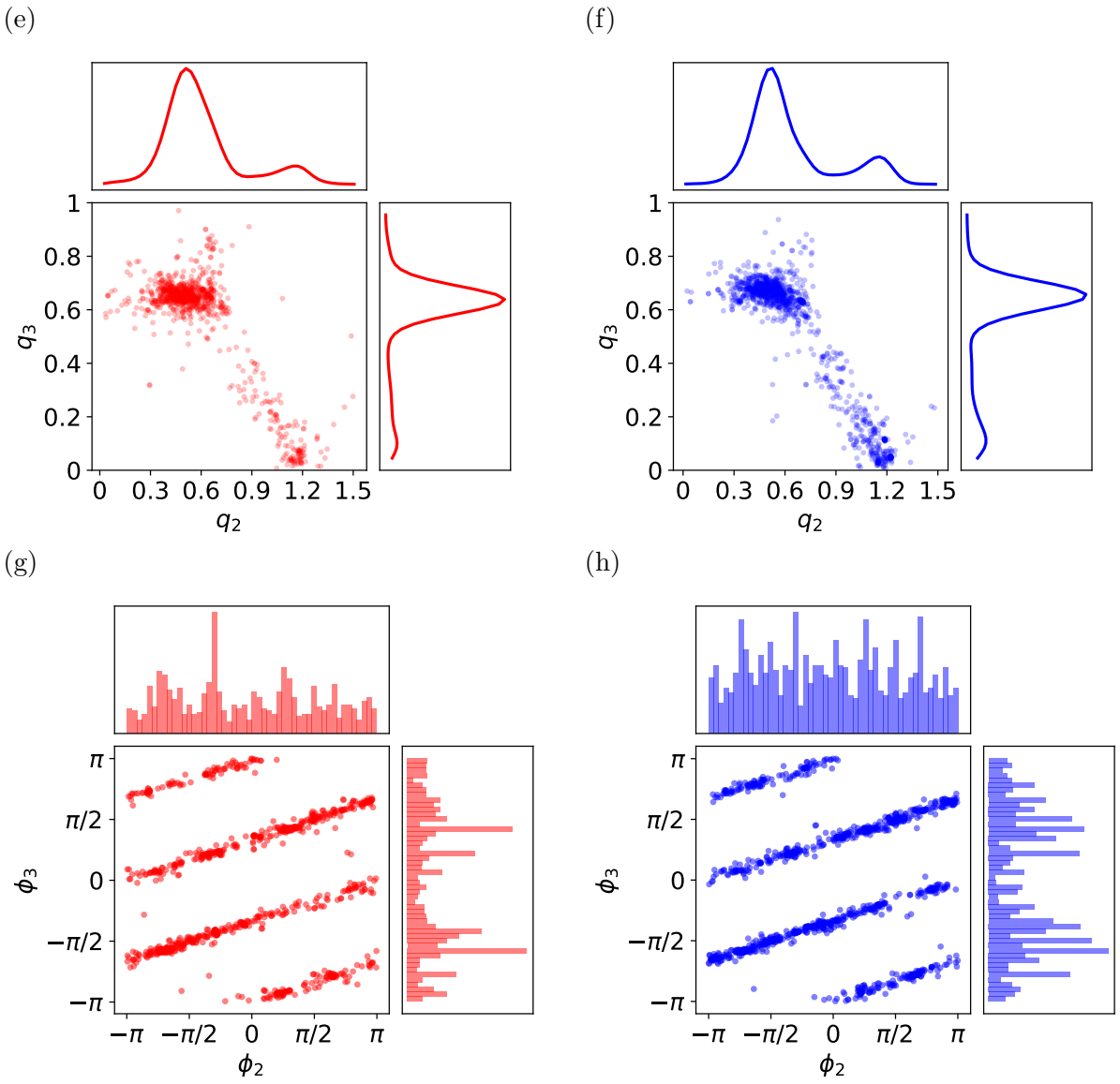


Figure S19: Continued

Acknowledgement

GMM thanks the EPSRC and MRC for financial support under grant number EP/L016044/1. GRH thanks the National Science Foundation (CHE-1800435) for support. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. The authors would like to acknowledge the University of Pittsburgh Center for Research Computing through the computational resources provided, and using resources provided by the Open Science Grid,[S17,S18] which is supported by the National Science Foundation award 1148698, and the U.S. Department of

Energy's Office of Science. L.C. thanks Susan Leung, Catherine Wong and Anne Nierobisch for helpful discussions.

References

(S1) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

(S2) Cremer, D. A General Definition of Ring Substituent Positions. *Isr. J. Chem.* **1980**, *20*, 12–19.

(S3) 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.

(S4) 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.

(S5) 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.

(S6) Landrum, G. RDKit: Open-source cheminformatics. `http://www.rdkit.org`.

(S7) 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.

(S8) 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.

(S9) 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.

(S10) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(S11) 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.

(S12) Flachsenberg, F.; Andresen, N.; Rarey, M. RingDecomposerLib: An Open-Source Implementation of Unique Ring Families and Other Cycle Bases. *J. Chem. Inf. Model.* **2017**, *57*, 122–126.

(S13) Dunbrack, R. L. Rotamer Libraries in the 21st Century. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431 – 440.

(S14) 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* **1993**, *49*, 910–928.

(S15) 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* **1994**, *50*, 395–404.

(S16) 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* **1996**, *52*, 882–891.

(S17) 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; p 012057.

(S18) Sfiligoi, I.; Bradley, D. C.; Holzman, B.; Mhashilkar, P.; Padhi, S.; Wurthwein, F. The Pilot Way to Grid Resources Using glideinWMS. 2009 WRI World Congr. Comput. Sci. Inf. Eng. 2009; pp 428–432.