Supplemental Information

BFMP: A method for discretizing and visualizing pyranose conformations

Spandana Makeneni[1], B. Lachele Foley[1], *, Robert J. Woods[1, 2,] *

[1]Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA, 30602, USA.

[2]School of Chemistry, National University of Ireland, Galway, University Road, Galway, Ireland.

*Authors to whom correspondence may be addressed.

Table S1: Comparison of Whitfield, Cremer-Pople and BFMP classification methods.

	Whitfield[a]	Cremer-Pople[b]		BFMP
		ϕ	θ	
1^1	0.017 $^4C_1$, 0.021 $^{1,4}B$ and 1.125 ^2S_O	271	91	$^5t_3(0.44)$
2^1	0.042 $^1C_4$, 0.827 $B_{1,4}$ and 0.046 $^OS_2$	178	96	$B_{1,4}(1.41)$
3^1	0.014 $^4C_1$, 0.001 $^{1,4}B$ and 1.204 ^2S_O	270	90	$^5t_3(0.34)$
4^2	1.012 $^4C_1$, 0.126 $B_{1,4}$ and 0.025 ^2S_O	191	158	$^4d_1(1.99)$
5^2	0.758 $^1C_4$, 0.377 $B_{1,4}$, 0.021 $^OS_2$ [c]0.754 E_4, 0.381 $^1C_4$ and 0.021 $^OS_2$	175	33	$^{3,5}t(0.82)$

[a]These values were obtained from the Table 4 in the Whitfield paper[3]. Structures 1-3, 4 & 5 correspond to 23-25, 30 & 31 from the same table.
[b]Values are in degrees
[c]In rows 1-4, only the "chair-boat-twistboat combination" is necessary to describe the conformations. However, here, these "intermediate conformations" were also necessary for a complete characterization. See the manuscript for details[3].
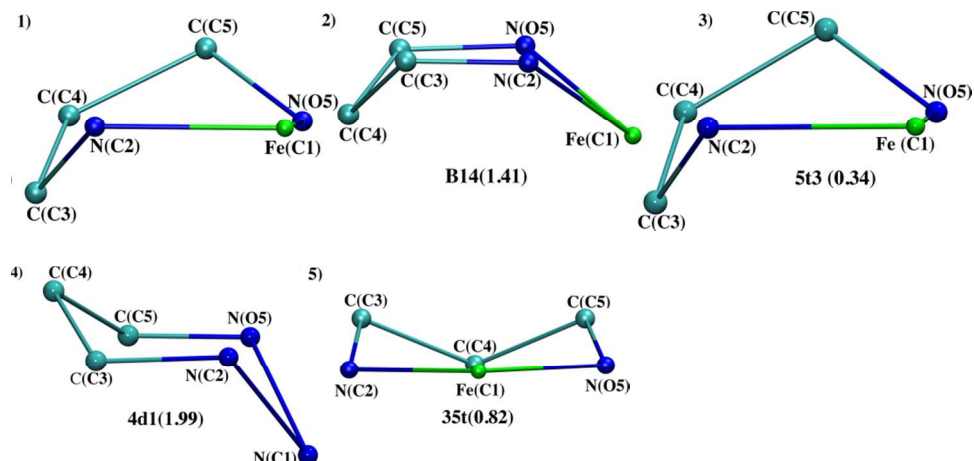
Figure S1: Ring conformations of the structures in table S1. The coordinates for these files were obtained from the SI of the Whitfield paper. To facilitate comparison of the Whitfield structures to BFMP naming as described in this manuscript, the original atom names in the PDB file (Fe,N,C) are shown with the corresponding pyranose atom names (C1, C2, C3, C4, C5 and O5) in parentheses. Ring 5 has been oriented to plainly show the best-fit plane.

Table S2: Reference planes for the 38 IUPAC conformations

Conformation	# Of reference planes	Atoms in the reference planes
$^1C_4$, $^4C_1$	3	C2 C3 C5 O5 C1 C3 C4 O5 C1 C2 C4 C5
^{14}B, B_{14}	3	C2 C3 C5 O5 C1 C4 C5 O5 C1 C2 C3 C4
$B_{3,O}$, $^{O,3}B$	3	C1 C2 C4 C5 C3 C4 C5 O5 C1 C2 C3 O5
$B_{2,5}$, $^{2,5}B$	3	C1 C3 C4 O5 C2 C3 C4 C5 C1 C2 C5 O5
$^3S_1$, $^1S_3$	2	C2 C4 C5 O5 C1 C2 C3 C5
$^OS_2$, ^2S_O	2	C1 C3 C4 C5 C1 C2 C4 O5
$^5S_1$, $^1S_5$	2	C2 C3 C4 O5 C1 C3 C5 O5
$^1H_2$, $^2H_1$	1	C3 C4 C5 O5
$^3H_2$, $^2H_3$	1	C1 C4 C5 O5
$^3H_4$, $^4H_3$	1	C1 C2 C5 O5
$^5H_4$, $^4H_5$	1	C1 C2 C3 O5
^5H_O, $^OH_5$	1	C1 C2 C3 C4

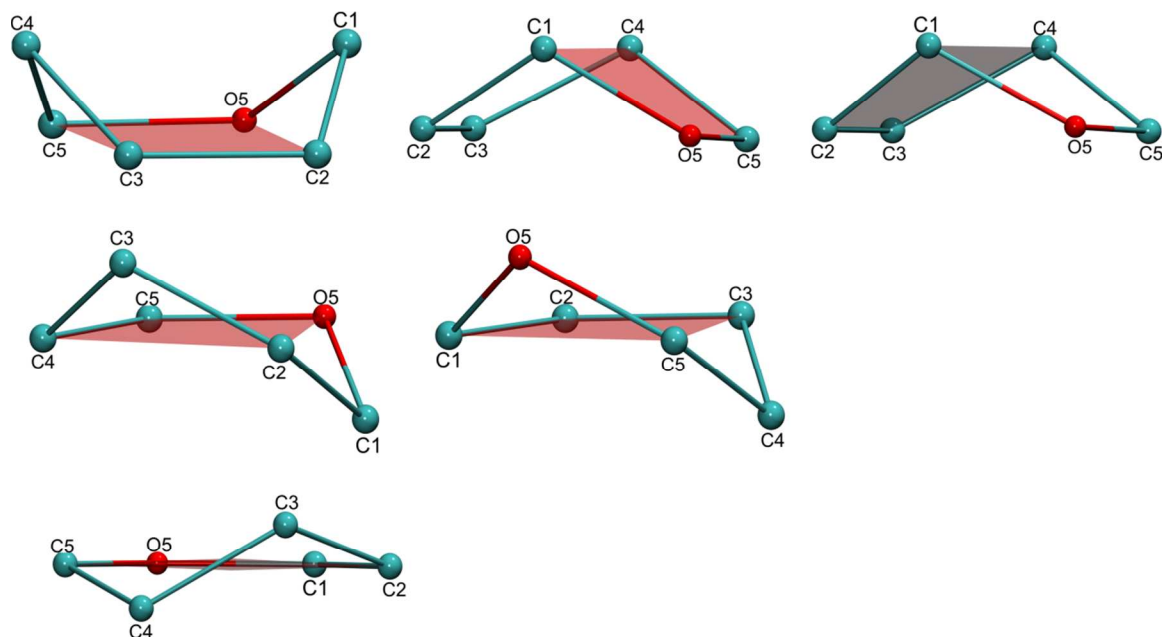|		
$^{1}H_{O}, {}^{O}H_{1}$	1	C2 C3 C4 C5
$^{1}E, E_{1}$	1	C2 C3 C4 C5 O5
$^{2}E, E_{2}$	1	C1 C3 C4 C5 O5
$^{3}E, E_{3}$	1	C1 C2 C4 C5 O5
$^{4}E, E_{4}$	1	C1 C2 C3 C5 O5
$^{5}E, E_{5}$	1	C1 C2 C3 C4 O5
$^{O}E, E_{O}$	1	C1 C2 C3 C4 C5

Figure S2: Top, center and below show reference planes that define idealized $^{1,4}B$, $^{3}S_{1}$ and $^{3}H_{4}$ conformations respectively.

Section S1 (Protocol for automated classification):

An automated version of the algorithm has been implemented in C. The version released in association with this publication is available for download at http://glycam.org/publication-materials/bfmp/. That version and any subsequent versions released should also be accessible from http://glycam.org/publication-materials.

Two detailed tutorials are also available. They illustrate portions of the calculations referenced in the main manuscript. "Determining the ring conformation of α–L-Idopyranose during the course of a 10ns MD simulation using BFMP" is available at http://glycam.org/docs/help/2014/07/28/bfmp-tutorial-1-3/, and "Determining the ring conformation of Iduronate from a PDB file using BFMP" is available at http://glycam.org/docs/help/2014/07/28/bfmp-tutorial-2/. Both tutorials, and any future versions, should be accessible via search at http://glycam.org/publication-materials.

The input, output, and the protocol used for the implementation are described in detail below.

Input

Coordinate input can come from a PDB file or from an AMBER format topology file with an associated trajectory or input-coordinate file. Additionally, a text input file containing the residue number and ring-atom names is required. The current version of the program can process only one residue at a time.

Output

The output file consists of three columns. The description for each column in the order that it appears in the file is given below.
1.) Simulation frame

Ring shapes of the residue number specified in the input file are calculated for each frame in a trajectory file. This column identifies the frame by index beginning with one. If the input file is a PDB file, the ring shape in the first model is determined.
2.) Standard/ Canonical Nomenclature

Where appropriate, the standard IUPAC nomenclature is listed in this column (See Protocol, below, for details)
3.) BFMP Nomenclature

The BFMP nomenclatures for all four-atom sets with dihedrals below the cutoff, along with the average dihedral angle, are listed in this column in order of increasing dihedral.

Here is example output from the program:

```
Timestep        Standard Nomenclature  Ring conformation
1                      -               3d6(1.278024)        5d2(9.817732)
```

The first line briefly describes the content of each column in the output and occurs once at the top of the file. The second line in this example indicates that in frame 1 (*Time step*) there are two best-fit four-membered planes below the cutoff value. The corresponding conformations are listed in the *Ring conformation* column. Conformations for all planes satisfying the cutoffs are listed in this column in the order of increasing average plane dihedral. The dash (-) in the *Standard Nomenclature* column indicates that the ring structure is too distorted to assign it a canonical descriptor. Note that the program merely numbers the atoms 1 through 6. In the manuscript, "6" has been changed to "O". The program can be used for any 6-membered ring (see also SI Table S1 and Figure S1).

Protocol
1.) <u>Extract the coordinates for the ring atoms from the coordinate file</u>.
2.) <u>Calculate best-fit planes for each set of four reference atoms in the ring</u>.
3.) <u>Determine which planes are possible reference planes</u>: Calculate dihedral angles around the perimeter of the quadrilateral described by each set of four reference atoms. Any set

of four atoms with an average dihedral angle less than the cutoff is considered to approximate a plane. The cutoff value can be adjusted based on user requirement. If the user does not specify a cutoff, the program will use the default cutoff (10°) that was used for calculations in the manuscript.

The cutoff was chosen based on examination of ring conformations from MD simulations using molecular visualization software. We observed that when the average dihedral angles are below 5° they appear to be very close to ideal symmetric conformations. The Cremer-Pople parameters for these structures are also unambiguously close to one of the ideal IUPAC conformations. We chose 10° as the cutoff because above this angle the structures begin to appear significantly distorted, and their CP parameters begin to enter ranges where assignment can be ambiguous. This is illustrated in Figure S3, which shows examples of a few ring conformations with various average torsion angles around the 10° cutoff.

4.) <u>Sort and rank the best-fit planes with average dihedral angles < the cutoff</u>: If there are no best-fit planes with average dihedral less than the cutoff, assign conformation "m".

5.) <u>If the conformation is not m, check if all six atoms in the ring are coplanar</u>: Dihedral angles are calculated for each set of four atoms around the ring perimeter starting from C1. A total of 6 dihedral angles can be calculated. The ring is classified as an "h" conformation if all six angles are less than 5°. This cutoff value, like the others, can be adjusted per system needs, but the program in its current form does not provide a convenient means for doing that.

6.) <u>If the ring shape is not classified as "m" or "h", check if there is a possible five-atom pseudo-plane (a "p" structure):</u> Consider a six-member ring with atoms labeled A, B, C, D, E and F. In this work, the ring shape is classified as "p", with atom D out of plane, if all of the following criteria are met (see Figure S4 for illustrations):

a. The differences between values for pairs of opposite dihedral angles satisfy all of these conditions:

$$\left|\,|\angle(EFAB)| - |\angle(FABC)|\,\right| < 5° \qquad (1)$$
$$\left|\,|\angle(ABCD)| - |\angle(DEFA)|\,\right| < 6° \qquad (2)$$
$$\left|\,|\angle(BCDE)| - |\angle(CDEF)|\,\right| < 9° \qquad (3)$$

b. The absolute values of dihedral angles $\angle(EFAB)$ and $\angle(FABC)$ are less than 12°.

These cutoff values were again chosen based on visual examination of the structures from MD simulations. We observed that at these cutoff values, the ring begins to transition into a half chair structure. This is illustrated in Figure S5. Currently, the program that is available for download does not provide a simple way for users to alter these values. Users with sufficient skills can edit the code to use new values and recompile the program.

7.) <u>If not m, h or p, check the top three best planes to see if the shape can be classified into one of the standard conformations.</u> It will be classified as an IUPAC conformation only if:

a. All reference planes for a given conformation (see Table S2) have average dihedrals less than the cutoff value specified by the user in step 3.

b. The positions of the out-of-plane atoms are consistent with the IUPAC descriptor.

8.) <u>Assign d, t and/or q conformations for all planes with average dihedrals below the cutoff value specified by the user in step 3.</u> Choose the conformation with the lowest average dihedral as best.

9.) <u>If yes from step 7, then the IUPAC conformation is listed under the standard nomenclature column</u>.

10.) <u>The four-atom BFMP conformations (d, t & q) identified for each best-fit plane in step 8 are listed in the third column ("*Ring conformation*") in order of increasing average dihedral</u>.
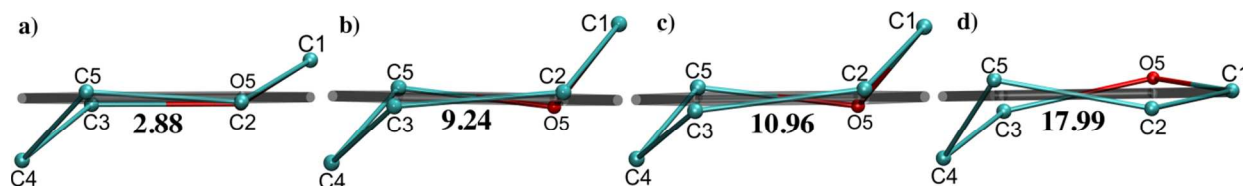

Figure S3: Ring conformations adopted by Iduronic acid during the course of an MD simulation. In each case, the gray line represents the average plane for atoms C2, C3, C5 and O5 and the average dihedral angles are shown below the plane. All values are in degrees. In a), the ring is in a 1C4 conformation, and the average dihedral for the indicated best-fit plane is 3°. Images b) and c) show two average dihedrals just above and below 10°. They also illustrate that increasing or decreasing the cutoff value by 1° might not make any significant difference to the results: the structures look nearly identical. However, note that near 10°, the atoms are beginning to stray noticeably from the plane. If this value is increased beyond 15°, then the atoms are significantly far from the plane. An example of a structure with an average dihedral angle of 18° is shown in figure d). Even if atom C1 in that structure were obviously above the average plane, the structure still could not be classified as a canonical chair conformation.
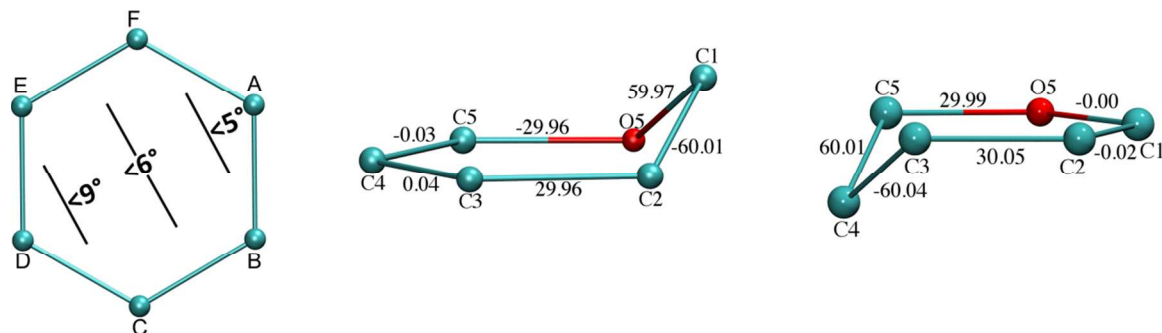

Figure S4: (Left) Six membered ring showing the difference requirement for pairs of opposite dihedral angles if the conformation is p with atom D out of plane. (Center) ^{1}p conformation (Right) p_4 conformation. Dihedral angles for the example p-type conformations are indicated (center and right).
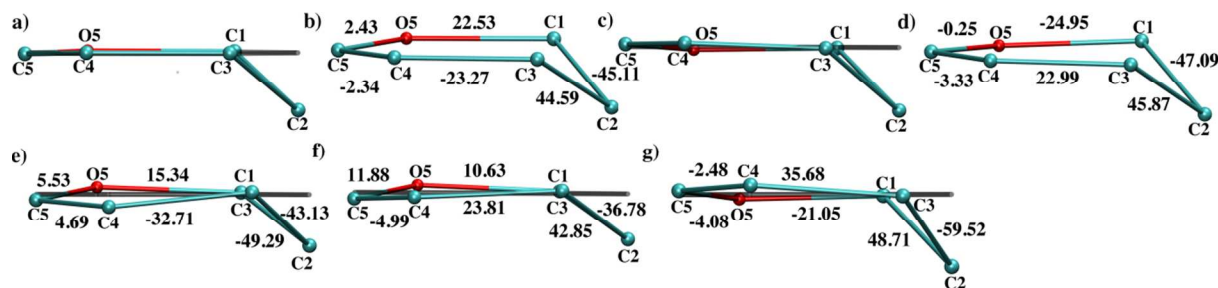
Figure S5: a) An ideal symmetric E2 conformation in which the differences, defined in 6a of the Protocol, above, in each of the opposite dihedral angles, shown in b), are < 1°. c) A structure that is typically classified as E2 by BFMP after employing the cutoff values. In c) the dihedral angles, shown in d), satisfy all conditions in 6a and 6b from the Protocol, above. In e) while requirements 6a(1) and 6a(3) are satisfied, 6a(2) is not, with the difference being >17°. Of the requirements in 6a, structures in f) and g) satisfy only (3) and (1), respectively. Note that all these structures satisfy the condition specified in step 6b. Also note that, in g), the atom C2 is sufficiently out of plane that it might be categorized as an envelope-type conformation. For simplicity, we chose not to include the effect on perception of the distances from the plane of the out of plane atoms. Instead, we focused only on the extent to which the positions of the other five atoms appeared to approximate a plane. Therefore, in the current work this is not classified as an envelope.

Table S3: List of Enzyme Substrate and inhibitor co-crystal complexes used for the inhibitor design analysis

Inhibitors	1HWW, 2WYI, 3BLB, 2WW0, 2WW2, 2OW6, 2OW7, 3EJP, 3EJQ, 3EJR, 3EJT, 3EJS, 3EJU, 1PS3, 1FO3, 2WVZ, 1KRF, 1G6I, 1HXK, 1FO2, 1KRE, 3QRY, 4AQ0, 4AD3, 4AD5, 1I75, 1OIM, 2J77, 2PWD, 2X2J, 2YA2, 2JKE, 3GBE, 3GXT, 3QFZ, 3QG0, 4IID, 2WZS, 2VMF, 3D4Y, 4AYQ, 4CD5, 4CD8, 4AYR, 2V3G, 2ALW, 2V38, 2J75, 3ZQ9
Substrates	1DL2, 1GW1, 1GVY, 1KKT, 1NXC, 1O7D, 1QWN, 1QWU, 1QX1, 1X9D, 2WBK, 2WHM, 2WW1, 2WW3, 3BUP, 3BUQ, 3BVT, 3BVU, 3BVV, 3BVW, BVX, 3CV5, 3CZN, 3PZI, 3PZO, 4AYP, 4JIE

Section S2 (MD Simulation Details):
MD simulations of α-L-Idopyranose were carried out using the AMBER12 software package[4]. Input files were constructued using the tleap[4] module of AmberTools12 with the GLYCAM_06h-1.dat parameters (http://glycam.org/params). The prep files used to build the

molecule are available for download at http://glycam.org/publication-materials. The molecule was solvated with the TIP3P water model in a cubic box with a 9 Å buffer. Energy minimization of 5000 steps of steepest decent followed by 5000 steps of conjugate gradient was performed in the NVT ensemble.The system was then heated to 300K for 50ps and equilibrated for 1ns, all using the NPT ensemble. Production MD simulations were run for 10ns with an integration step size of 1fs. Snapshots were collected at every picosecond for subsequent analysis. All simulation input and output files are available at http://glycam.org/publication-materials.
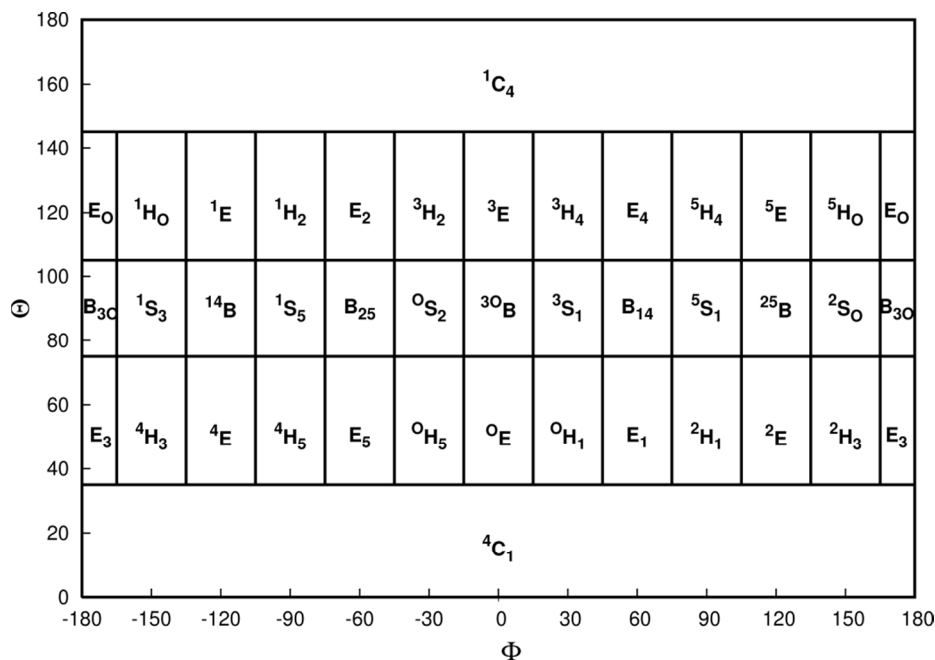

Figure S6 Mappings used to classify the standard IUPAC conformations based on Cremer-Pople puckering parameters. For convenience in interpreting these data sets, values in ϕ are reported from -180 to 180 rather than the standard 0 to 360. Also, ranges for H and E type conformations reflect an average median position (their median positions in θ are offset by 15°).

Table S4: Non-standard Conformations identified by the BFMP method during the course of a 10ns MD simulation of α-L-Idopyranose (These correspond to the conformations on the y-axis in Figure 6 of the paper)

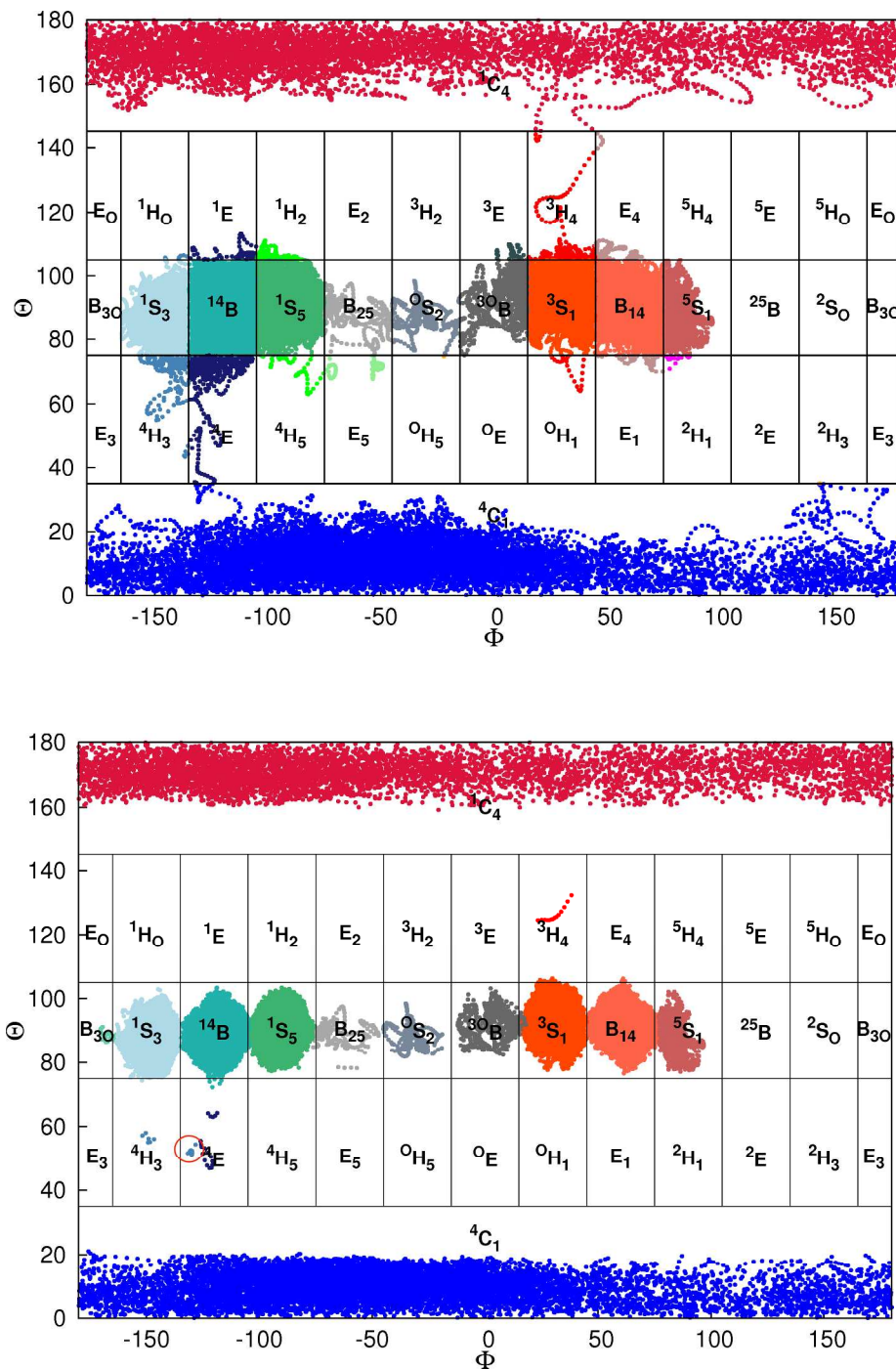| Conformation type	Conformations
d	$^3d_6$, $^5d_2$, $^1d_4$, d_{14}, ^{36}d, d_{25}, ^{14}d, $^4d_1$, $^2d_5$, $^6d_3$
t	t_{24}, ^{35}t, $^6t_4$, $^3t_1$, $^2t_4$, $^5t_1$, $^3t_5$, $^4t_2$, $^1t_5$, $^1t_3$, $^4t_6$, ^{24}t, t_{35}
q	^{23}q, q_{45}, ^{56}q, q_{12}, ^{34}q, q_{56}, q_{23}, ^{45}q, ^{16}q, ^{12}q

Figure S7: See also Figures S6 and S8. (Top) Conformations identified by standard mappings of IUPAC nomenclature to the Cremer-Pople parameters for the data from Figure S6. The colors correspond to conformations as defined by the mapping grid. (Bottom) Canonical IUPAC conformations, identified by the BFMP method, colored as in the upper figure. The data from the upper figure that are missing in bottom figure represent conformations that are not easily mapped to IUPAC conformations. Where colors are present outside their grid lines, those standard

conformations were identified in regions outside the grid-based mapping. The red circle in the region labeled ^{4}E specifies the location of the structure featured in Figure S8.
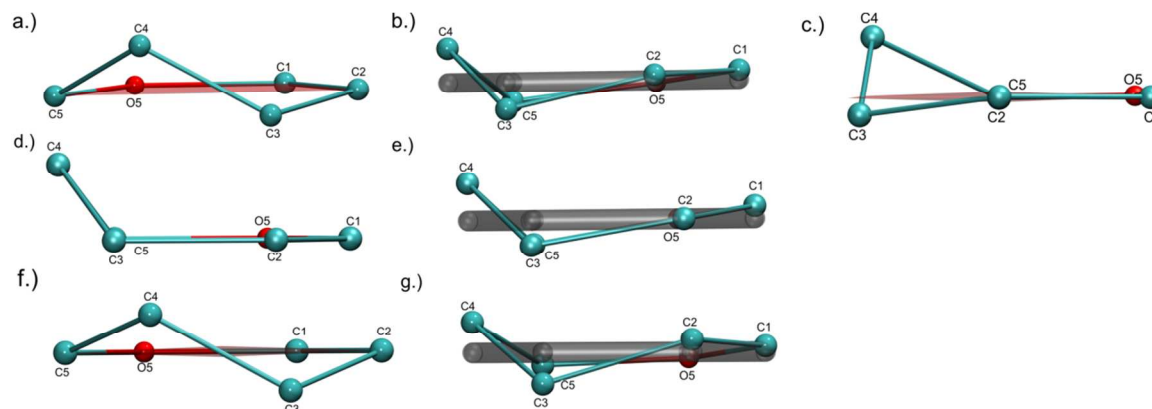

Figure S8: a) This structure is identified by BFMP as ^{4}H$_3$, but would normally be mapped to ^{4}E from Cremer-Pople parameters (its CP parameters are marked with a red circle in Figure S7). a) While it is a distorted ^{4}H$_3$, c) it does not make a convincing envelope since the atoms C1, C2, C3, C5, and O5 are not very coplanar. The reason for the mapping can be easily illustrated (b, d-g). Images d) & f) show idealized ^{4}E and ^{4}H$_3$ conformations, respectively, in a typical IUPAC orientation. Images e) & g) show a cross-section of the CP average 6-atom plane, illustrated in gray, for d) and f). Note the relative positions of the atoms with respect to the CP 6-atom plane. Image b.) shows the 6-atom plane from a) as in e) & g). Note that while atoms C2 and O5 are slightly above and below the plane (like g), they are very near the average plane (like e). The relative positions of C3 and C5 in b) are also partway between e) and g). Of course, the decision to classify a structure such as the one in a) as H or E depends on the judgment of the observer, but these differences can complicate mapping between the two methods when conformations are between classifications.

References:

1. Herron, N.; Grzybowski, J. J.; Matsumoto, N.; Zimmer, L. L.; Christoph, G. G.; Busch, D. H., A new family of clathrochelate complexes produced by a ligand rearrangement reaction. *J. Am. Chem. Soc.* **1982,** *104*, 1999-2007.

2. Dewar, R.; Fleischer, E., Structure of a Reduced Nickel(II) Tetradentate Macrocycle. *Nature* **1969,** *222*, 372-373.

3. Bérces, A.; Whitfield, D. M.; Nukada, T., Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature. *Tetrahedron* **2001,** *57*, 477-491.

4. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 12. University of California, San Francisco: 2012.