### Is it usual to be unusual? An investigation into

### molecular conformations in organic crystals

Sarah E. Wright<sup>1</sup>, Mathew J. Bryant<sup>2</sup> and Aurora J. Cruz-Cabeza<sup>1,3</sup>\*

1 School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, UK

2 The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK

3 Chemical Development, Pharmaceutical Technology & Development, AstraZeneca, Macclesfield

Campus, SK10 2NA

\*email: aurora.cruzcabeza@manchester.ac.uk

1.	Mogul Definition	. 2
2.	Choice of Torsion Fragment to Describe a Rotatable Bond	. 2
3.	Intramolecular Hydrogen Bond Detection	.3
4. E	xample where a torsion fragment has no preferential distribution (the	
dist	tribution is relatively flat)	.3
5.	Example of unusual, from current Mogul definition where the observed	
valı	ue is at a potential energy minima	.3
6.	Example of a distribution where the torsion fragment in the query molecu	le
is c	onstrained by an intramolecular hydrogen bond but much of the rest of the	)
dist	ribution is not	.4
7.	Example of Torsion Analysis with Mogul	. 5
8.	Statistics of Unusual Torsions in the CSD (New Definition)	.6
9.	Comparing Usual and Unusual: Geometry	.9
10.	Histograms for P <sub>adjust</sub> and P <sub>change</sub>	10
11.	Data figures before new filters were applied	11

#### 1. Mogul Definition

The Mogul definition for an unusual torsion is that less than 5% of the distribution is

within 10° of the observed value. This has been visualised in

Figure S1.



Figure S1: Schematic to support the Mogul definition of unusual.

#### 2. Choice of Torsion Fragment to Describe a Rotatable Bond

Each rotatable bond can be defined by a torsion angle generated by the arbitrary choice of four consecutive atoms (the bond of interest being in the middle). There are various possible choices of atoms which result in various different torsions which define the same rotatable bond. For our analysis, our searches for each rotatable bond were done using the all possible torsion definitions and we then kept the torsion definition affording the largest number of CSD hits. This decision was made with the assumption that the distribution with more hits would be statistically more significant, thus would afford more reliable statistical results. If multiple distributions contained the same number of hits and if they agreed that the torsion angle was unusual, then it was classified as unusual. However if the distribution classifications disagreed, a "usual" classification was given prevalence over the "unusual" one.

A	С	Torsion Fragment	No. Hits	Classificaiton	
$\sim$	,	A-X-X-C	1000	Unusual	
		A-X-X-D	1000	Usual	
В	D	B-X-X-C	900		
		B-X-X-D	800		

Figure S2: Schematic to support the choice of torsion fragment to describe a rotatable bond. The chosen torsion fragment is in red.

#### 3. Intramolecular Hydrogen Bond Detection

The molecule was searched for intramolecular hydrogen bonds, using the criteria in Table S1 and the shortest bonded path between the hydrogen bonded atoms. If central bond of the torsion angle was in this path then it was flagged as being constrained.

 Table S1:
 Criteria applied to define a hydrogen bond using the CSD python API.

Hydrogen Bond Criteria					
intermolecular = 'intramolecular'					
path_length_range = (-1, 999)					
$angle_tolerance = 120.0$					
vdw_corrected = False					
distance_range = $(0,2.8)$					
require_hydrogens = True					

## 4. Example where a torsion fragment has no preferential distribution (the distribution is relatively flat).

Difficulties were found when CSD torsion angle does not have a preferential distribution, and thus is flat. An example of this is the C10\_C1\_C19\_C20 fragment of 3'-amino-2-(4-bromophenyl)-1H,1'H-2,2'-bi-indene-1,1',3(2H)-trione, where the distribution does not show a preference for any torsion angle.



Figure S3: The C10\_C1\_C19\_C20 fragment (indicated by balls and atomic labelling) of ENUNIL returns a non-preferential distribution, which is relatively flat.

# 5. Example of unusual, from current Mogul definition where the observed value is at a potential energy minima.

An individual analysis of some of the unusual torsions identified with the current Mogul definition revealed drawbacks of this broad definition. This is illustrated with one example in Figure S4. Here we analyse the outcome of the analysis in Nitrofural (WERVEU) where the observed torsion is found close to 180°. This observed torsion is flagged as unusual. Whilst this value may be unusual with regard to change (its in a local minimum), the conformation observed in the crystal is only slightly adjusted from its closest local minimum located at 180°.



Figure S4: CSD distribution (green bins) and PES (red line) for compound WERVEU in the CSD. The original Mogul definition classifies the torsion as unusual. Whilst this may be unusual change, this conformation is usual with regard to adjustment. The atoms involved in the torsion angle of interest are shown as spheres.

6. Example of a distribution where the torsion fragment in the query molecule is constrained by an intramolecular hydrogen bond but much of the rest of the distribution is not.



Figure S5: CSD distribution for the O1\_C11\_C8\_C9 torsion fragment of 1,8-Dihydroxy-2naphthaldehyde, with an observed value of 179° shown by the black line. The distribution indicated in yellow, shows only the hits whereby the torsion angle is held by an intramolecular hydrogen bond, like the query structure. The structure of 1,8-Dihydroxy-2-naphthaldehyde is overlain with the torsion angle indicated by an arrow and the atoms involved in the torsion angle of interest are shown as spheres.

#### 7. Example of Torsion Analysis with Mogul

Here we show the Mogul torsional analysis for the compound N-phenyl-2-(pyridin-2-ylmethylene)hydrazinecarboxamide with refcode EMIPEW02. This compound has five rotatable bonds, however the Mogul search returns 10 torsion fragments, whereby two are found to be unusual. Both of these unusual torsion angles are around the C8-C9 rotatable bond.

R bond	<b>Torsion Fragment</b>	Value	N hits	Local Density	Classification
C7-N1	O1 C7 N1 C1	-2.684	5395	0.886	Not unusual (enough hits)
C7-N1	N4 C7 N1 C1	177.144	104	0.942	Not unusual (enough hits)
C1-N1	C2 C1 N1 C7	-21.262	7620	0.202	Not unusual (enough hits)
C1-N1	C6 C1 N1 C7	159.82	7620	0.213	Not unusual (enough hits)
C8-N2	C9 C8 N2 N4	179.737	282	0.869	Not unusual (enough hits)
C7-N4	N1 C7 N4 N2	-9.273	101	0.901	Not unusual (enough hits)
C7-N4	O1 C7 N4 N2	170.564	183	0.661	Not unusual (enough hits)
N4-N2	C7 N4 N2 C8	174.668	187	0.973	Not unusual (enough hits)
C8-C9	C10 C9 C8 N2	-28.533	376	0.032	Unusual (enough hits)
C8-C9	N2 C8 C9 N3	148.502	382	0.016	Unusual (enough hits)

#### EMIPEW02

### Figure S6: Output from Mogul for N-phenyl-2-(pyridin-2-ylmethylene)hydrazinecarboxamide (EMIPEW02).

#### 8. Statistics of Unusual Torsions in the CSD (New Definition)

### Statistics of Unusual Torsions divided into simple and non-simple organic molecules.

The flexibility of a molecule was used to define whether it is simple or non-simple, where compounds with 4 or more rotatable bonds are defined as non-simple. There is little difference in the ratio of torsions being classified as unusual adjustment to unusual change and to unusual both. However for non-simple organic molecules the percent of structures with at least one unusual torsion angle doubles (18% to 44%). There is also a higher prevalence of structures having multiple unusual torsions in different categories. In particular, the number of structures with both a torsion angle with unusual adjustment and a torsion angle with unusual change increases by 10%.



Figure S7: Venn diagram showing the composition of unusual structures in the single component subset with fewer than four rotatable bonds by the different types of unusual torsion identified (14 175 unusual crystal structures).



Figure S8: Venn diagram showing the composition of unusual structures in the single component subset with more than or equal to four rotatable bonds by the different types of unusual torsion identified (30 133 unusual crystal structures).

Venn diagrams for other solid form types (solvates, hydrates, cocrystals and salts)



Figure S9: Venn diagram showing the composition of unusual structures in the solvate subset by the different types of unusual torsion identified (2 934 unusual crystal structures).



Figure S10: Venn diagram showing the composition of unusual structures in the hydrate subset by the different types of unusual torsion identified (2 002 unusual crystal structures).



Figure S11: Venn diagram showing the composition of unusual structures in the cocrystal subset by the different types of unusual torsion identified (1 301 unusual crystal structures).



Figure S12: Venn diagram showing the composition of unusual structures in the salt subset by the different types of unusual torsion identified (3 750 unusual crystal structures).

### Example output for WEBFOY, a structure containing torsions with all four classifications.

R bond	Torsion Fragment	Value	N hits	P <sub>adjust</sub>	P <sub>change</sub>	Classification	Comment
C29-O1	C28_O1_C29_C34	26.95	3711	0.008	0.995	Unusual Adjustment	
O1-C28	C29_O1_C28_C27	70.40	5522	0.063	0.13	Unusual Change	
C28-C27	O1_C28_C27_C26	66.02	5131	0.277	0.808	Usual	
C27-C26	C25_C26_C27_C28	-157.86	141	0.009	1.0	Unusual Adjustment	Few hits after filtering
C26-C25	C27_C26_C25_N1	-41.74	3174	0.009	0.372	Unusual Both	
C25-N1	C26_C25_N1_C5	-72.77	579	0.088	1.0	Usual	
C7-C23	C6_C7_C23_C24	179.88	129				Few hits

Table S2: Output for WEBFOY, an example structure with all four torsion classifications.

#### 9. Comparing Usual and Unusual: Geometry

Figure S13 shows a scatter plot of adjustment energies versus torsion adjustment. From this figure we can infer that geometrical deviation does not correlate with adjustment energy. Each torsion angle may have a different stiffness and thus a small torsion adjustment may result in a high adjustment energy whilst the same variation in a different torsion may just require little energy. As such, there are no correlations between energies and torsional adjustments since one would need to consider torsions of individual nature separately. Most adjustments occur between 20 to  $50^{\circ}$ .



Figure S13: The adjustment energies calculated against the adjustment of the torsion angle from the centre of the peak for unusually adjusted torsions (left) and usually adjusted torsions (right).

#### 10. Histograms for Padjust and Pchange

The distribution of  $P_{adust}$  for all analysed torsions is almost Boltzmann like with the majority of torsion angles having a  $P_{adjust}$  value less than 0.3, as can be seen in Figure S14. For a torsion angle to not be observed as in the most popular peak is incredibly rare, with less than 10% of torsion angles not being in the most popular peak, with a  $P_{change}$  less than 0.95, as can be seen in Figure S15.



Figure S14: Histograms of the P<sub>adjust</sub> for all analysed torsion angles, with a bin size of 0.05.



Figure S15: Histograms of the P<sub>change</sub> for all analysed torsion angles, with a bin size of 0.05.

#### 11. Data figures before new filters were applied

We found that the CSD distributions retrieved by MOGUL were not representative of the query torsion in some cases. Two factors were found extremely relevant in this analysis a) whether the rotatable bond was cyclic or acyclic and b) whether the rotatable bond was constrained by an intramolecular hydrogen bond. If a torsion was derived to be unusual according to change, we then applied our own cyclicity and intramolecular hydrogen bond filters to further evaluate the torsion behaviour. The data **before** these filters were applied is presented in the following sections.

#### Defining cut-off parameters for Adjustment and Change



Figure S16: Scatter plot of adjustment energies versus P<sub>adjust</sub> before cyclic hits and intramolecular hydrogen bonds were accounted for. The green line shows the chosen cut-off for

unusual adjustment and the red line shows RT at room temperature (2.5 kJ mol<sup>-1</sup>). The inner graph magnifies the plot at lower values of P<sub>adjust</sub>.



Figure S17: Scatter plot of change energies versus P<sub>change</sub> before cyclic hits and intramolecular hydrogen bonds were accounted for. The green line shows the chosen cut-off for unusual adjustment and the red lines shows RT at room temperature (±2.5 kJ mol<sup>-1</sup>).

Venn diagrams for other solid form types (solvates, hydrates, cocrystals and salts)



Figure S18: Venn diagram for showing the classification of unusual structures (46 115 crystal structures).



Figure S19: Venn diagram showing the composition of unusual structures in the solvate subset by the different types of unusual torsion identified (3 025 unusual crystal structures).



Figure S20: Venn diagram showing the composition of unusual structures in the hydrate subset by the different types of unusual torsion identified (2 108 unusual crystal structures).



Figure S21: Venn diagram showing the composition of unusual structures in the cocrystal subset by the different types of unusual torsion identified (1 365 unusual crystal structures).



Figure S22: Venn diagram showing the composition of unusual structures in the salt subset by the different types of unusual torsion identified (3 867 unusual crystal structures).



#### **Comparing Usual and Unusual: Energetics**

Figure S23: Histograms of adjustment/change energies calculated for torsions with a) unusual adjustment, b) usual adjustment, c) unusual change and d) usual change. Bin size of 1 kJ/mol.

#### **Comparing Usual and Unusual: Geometry**



Figure S24: Histograms of adjustment/change of the torsion angle with a) unusual adjustment, b) usual adjustment, c) unusual change and d) usual change. Bin size of 5°.



Figure S25: The adjustment energies calculated against the adjustment of the torsion angle from the centre of the peak for unusually adjusted torsions (left) and usually adjusted torsions (right).