JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

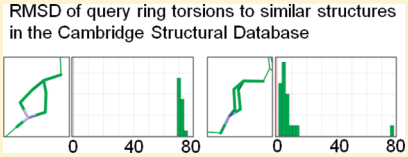
pubs.acs.org/jcim

Validating and Understanding Ring Conformations Using Small Molecule Crystallographic Data

Simon J. Cottrell,* Tjelvar S. G. Olsson, Robin Taylor,† Jason C. Cole, and John W. Liebeschuetz

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, U.K.

ABSTRACT: Understanding the conformational preferences of ring structures is fundamental to structure-based drug design. Although the Cambridge Structural Database (CSD) provides information on the preferred conformations of small molecules, analyzing this data can be very time-consuming. In order to overcome this hurdle, tools have been developed for quickly extracting geometrical preferences from the CSD. Here we describe how the program Mogul has been extended to analyze and compare ring conformations, using a library derived from over 900 000 ring fragments in the CSD. We illustrate how these can be used to understand the conformational preferences of molecules in a crystal lattice and bound to proteins.



INTRODUCTION

Ring systems are typically the scaffolds around which drug molecules are built.[1,2] The key role of rings in drug compounds is clearly illustrated by their use in the description and exploration of drug and chemical space,[3−6] as well as the existence of ring-specific databases as aids for medicinal chemists,[7,8] Where these rings are not rigid, it is clearly important to understand their conformational preferences.[9−12]

One way of gaining an understanding of the conformational preferences of rings is to use molecular modeling. The idea of modeling ring conformational space was pioneered by the work of Kilpatrick, Pitzer, and Spitzer,[13] who used the opposing torsional and angular forces of cyclopentane to model its structure. In automatic 2D to 3D model building, the requirement for ring closure means that atoms in rings are often treated differently to their straight chain equivalents. A common approach to dealing with rings in 2D to 3D conversion is to sample sets of stored ring templates,[14] although methods for sampling ring conformations dynamically have also been described.[15,16] Modern 2D to 3D software packages tend to be able to generate ensembles of ring conformations. The quality of these 2D to 3D software packages is generally evaluated using structures taken from the Cambridge Structural Database (CSD).[17,18] Not all publications on 3D conformer generation give much information on the number of ring templates used or how the templates were generated. One example that does is the recent publication on ConfGen.[19] ConfGen generates ring templates using Monte Carlo torsional sampling in conjunction with molecular mechanics minimization. High-energy conformations are excluded by removing entries with free energies 50 kJ/mol greater than the lowest energy conformation. Conformational diversity is obtained by making sure that no pairs of conformers have a root-mean-square deviation (rmsd) of atomic positions less than 0.25 Å. The ConfGen ring methodology clearly illustrates the nontrivial nature of modeling ring conformations. Force field based methodologies such as this can also suffer from insufficiently well derived parameters for certain combinations of atom types. In addition there is always the problem of ensuring that enough of the search space is covered to make it likely that the lowest energy conformer is found.

A complementary method to molecular modeling is to use experimental data. While the CSD[20] is recognized as a high-quality source of information on low-energy conformations of small molecules,[21] extracting and analyzing this information can be very time-consuming. Take, for example, the classic study by Engh and Huber,[22] which used data from the CSD to create restraints for use in protein structure refinement. To retrieve the appropriate fragments from the CSD, it was necessary to set up several nontrivial searches manually. In order to tackle this problem, methods for automatically retrieving precalculated information on all bond lengths, valence angles, and torsion angles, and comparing them to a query molecule, have been developed, such as the program Mogul,[23] which forms part of the Cambridge Structural Database System (CSDS).

Methods for automatically identifying and comparing ring structures have so far been lacking from Mogul. This partly stems from the fact that describing and comparing ring structures is more complicated than describing and comparing simple geometric features, such as bond lengths and torsion angles, which can be unambiguously represented by single values. The analysis of ring structures is further complicated by the concepts of pseudorotation[13] and symmetry. One metric for comparing rings is the Cremer−Pople ring puckering parameters;[24] however, interpreting these is complicated as the values lie in an $(n-3)$-dimensional space (where n is the size of the ring).

In this paper, we describe a new methodology for analyzing and comparing ring conformations.

METHODOLOGY

Software Overview. Mogul is a program that allows rapid access to CSD-derived information on bond lengths, valence angles, and torsion angles. This is achieved by arranging the

Received: September 15, 2011
Published: February 28, 2012

   dx.doi.org/10.1021/ci200439d | J. Chem. Inf. Model. 2012, 52, 956−962

fragments from the CSD in a tree structure, where each branch point represents an aspect of the chemical environment, encoded by a descriptor known as a key. Each fragment type has its own set of keys and tree structure. The leaf nodes of the tree consist of a set of fragments with identical values of every key. Thus, the levels closest to the root represent the most fundamental classification of the fragments and those closest to the leaves represent the most fine-grained classification.

To perform a search, key values for the query fragment are calculated and subsequently matched to the tree. If the tree contains a leaf with identical key values, the fragments on that leaf are retrieved and classified as "exact hits". The search can be generalized if there are insufficient hits, as specified by a user-defined threshold. The user thus obtains a collection of hit fragments whose chemical environment is as similar as possible to that of the query.

The distribution of the geometric value of the hit fragments is displayed as a histogram, and the hits can be browsed using a built-in visualizer. The query may be presented as a 2D or 3D molecule with the fragment of interest indicated. It can thus be ascertained whether the geometry of the query fragment is typical of chemically similar fragments in the CSD and geometric features can be classified as either usual or unusual based upon their similarity to entries in the CSD.

Ring Keys. The rings within the CSD were classified using a three-level tree. The keys used to classify the rings at each level were the following: (1) the number of atoms in the ring; (2) the set of atom types[25] of each atom in the ring; (3) a classification of the substituents on each ring atom according to their orientation and bulk. A canonicalization procedure, which generates key values for every possible atom ordering and uses a lexicographical comparison to pick the one with the "lowest" values, was applied to the atom ordering to ensure that two equivalent rings were not classified differently due to arbitrarily selecting different starting atoms or direction of ordering. Fused and bridged ring systems were excluded.

The orientation of each substituent was evaluated with respect to the plane defined by the two adjacent ring bonds. If the substituent atom adjacent to the ring was within 0.001 Å of this plane, it was considered to be "in plane"; otherwise it was considered to be "above" or "below" the plane, where "above" was arbitrarily defined as the direction from which the ring atoms appear to be ordered clockwise. For rings bearing two or more substituents, the orientation information effectively defines the relative stereochemistry of the substituents, e.g. would distinguish a ribose ring from an arabinose ring.

The bulk of a substituent was classified as either "hydrogen", "primary", or "bulky". The "hydrogen" classification obviously applied when the substituent was just a hydrogen atom. Groups with no more than one non-hydrogen substituent on the first atom, e.g. CH_3-, NH_2-, $CHNH-$, $(CH_3)_3CCH_2-$, were classified as "primary". Groups with two or more non-hydrogen substituents on this atom, e.g. $(CH_3)_2CH-$, NO_2-, were classified as "bulky". Thus, a substituent may in fact contain a long chain, but still be classified as "primary", providing there was no branching at the atom adjacent to the ring.

It was assumed that the absolute stereochemistry of the rings was not significant. There were two reasons for this. If the compound crystallized in a space group with an inversion operator, then the unit cell would contain two configurations of the ring (mirror images of each other). However, in most cases, only one of these would be stored in the database. Second, even if the space group did not contain an inversion operator, the absolute stereochemistry of the stored structure would not always be accurate, i.e. the actual structure could be a mirror image of that stored in the database. Therefore, when evaluating the key for level three, the inverted version of the ring was also considered. If two rings could obtain the same level three key value by inverting one of the rings they were placed in the same branch of the tree.

The ring tree library used in this study contains 941 863 fragments, derived from 525 095 entries (CSD version 5.32, November 2010).

Comparing the Geometry of Rings. In our geometric comparison of ring conformations we only considered variations in torsion angles. The variation that is observed for bond lengths and bond angles, among fragments in a similar chemical environment, is generally much smaller than that observed for torsion angles, at least for bonds considered to be "rotatable". Therefore, when considering the conformation of a ring, it is reasonable to assume that the bond lengths and bond angles within the ring are fixed a priori and that any possible variation in the conformation of the ring is due to variation in the torsions around the rotatable bonds of the ring. Rings consisting entirely of nonrotatable bonds show almost no variation in conformation.

The geometry of the three original fragment types (bond, angle, and torsion) can each be described by a single scalar value (bond length, bond angle, and torsion angle, respectively). Thus, for a given set of fragments, one can plot a histogram of the geometry distribution of those fragments.

It is not possible to describe the geometry of a ring in the same way as a bond length or a valence angle, with a single value. Assuming that only the torsion angles can vary then the conformation of an n-membered ring has $n - 3$ degrees of freedom. There are n torsions which can each vary, but they are constrained by the fact that they must form a continuous ring. The selection of appropriate parameters to describe these degrees of freedom has been studied extensively, most notably by Cremer and Pople[24] who devised a general definition of ring puckering. The Cremer–Pople ring puckering definition is based on the generation of a unique mean plane from which the ring puckering is described in terms of amplitude and phase coordinates. Several applications have used the Cremer–Pople parameters, for example, the widely used WHAT_CHECK[26] validation suite. In the early 1990s this puckering definition was also used to investigate the preferred ring conformations of medium sized-rings in the CSD.[10−12] However, interpreting the results is difficult. We have therefore developed a scalar pairwise distance measure for comparing two ring conformations, which we will refer to as the torsion rmsd (described below).

The calculation of the torsion rmsd between two rings requires a one-to-one mapping of consecutive bonds in the two rings, which implies a corresponding mapping of the atoms. Since in a search, we only generalize up to the level where all rings have the same atom types, the atom type matching is used to provide the mapping of the atoms and, hence, of the bonds. If either of the rings contains topological symmetry in terms of its atom types, then the rmsd for all possible mappings is calculated, and the lowest value is taken as the rmsd between the two rings.

(1) For a given mapping, the rmsd is calculated as follows: For each bond in the two rings, the intracyclic torsion angle is calculated. For each pair of mapped bonds, the difference between the two torsions in calculated.

dx.doi.org/10.1021/ci200439d | J. Chem. Inf. Model. 2012, 52, 956−962

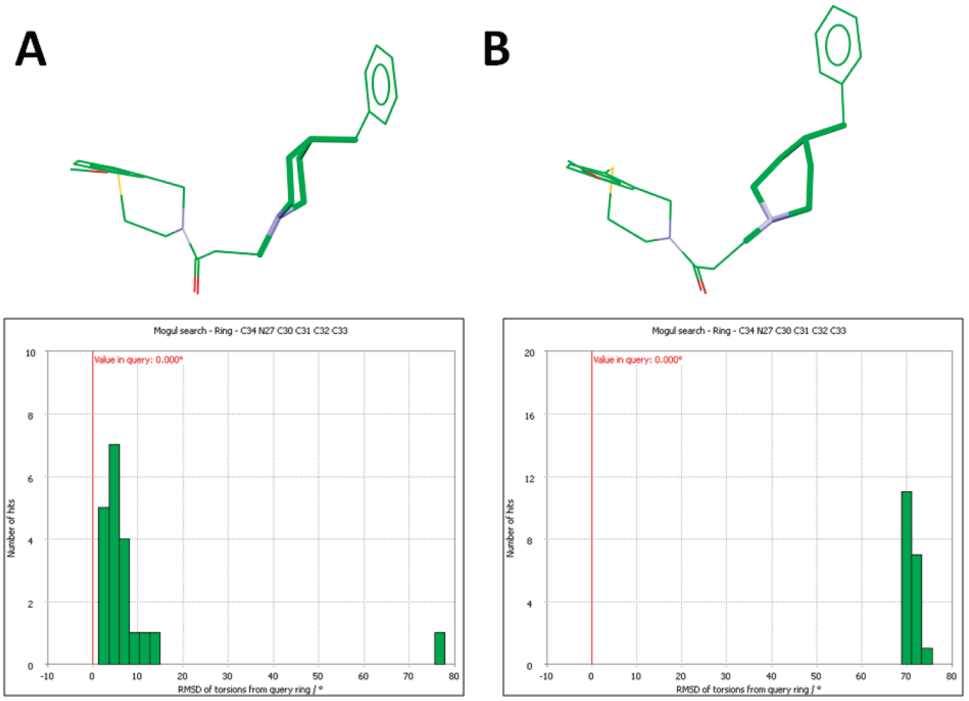


Figure 1. Two ring conformations of piperidine in PDB entry code 1hak. The chain A ring (left) is in a chair conformation whereas that of chain B (right) is in a twist-boat conformation. The histograms displaying the distribution of the rmsd with respect to the data in the CSD clearly show that the chair conformation is more common. It is worth noting that the same hits are not retrieved for these two queries. The reason for this is that Mogul classifies the substituents on the N and C4 atoms as being cis in the chair ring conformer but trans in the twist-boat conformer. Because the relative stereochemistry of the substituents is used by Mogul to describe the chemical environment of the rings, these two rings are not considered to be identical. Since the nitrogen atom is presumably capable of inversion, there is an argument for ignoring the orientation of the N substituent when generating the keys, in which case both rings would find the same (merged) set of hit fragments; but, the results would still lead to the conclusion that the twist-boat conformer is unusual and the chair conformer is not.

Theoretically, this takes a value in the range 0−180°; however, in practice, it can only take a value much less than 180° except for very large rings.

(2) The root-mean-square of all such differences around the ring is calculated, giving the torsion rmsd for the ring.

(3) We make the assumption that the actual structure of the rings may be the mirror image of that stored in the database. Therefore, the rmsd with one of the rings in an inverted conformation is also calculated; the lesser value is taken as the rmsd for the mapping.

The geometry of a ring fragment is compared to the hits from the library by plotting the distribution of the torsion rmsd. This has the advantage that the user can readily assess the geometry of the query fragment in relation to the data in the CSD. The torsion rmsd metric can be used to classify rings as "usual" or "unusual". Ring conformations are classified as unusual if less than 5% of the experimental values are within 10° of the conformation in the query.

Ring Clustering. A disadvantage of the torsion rmsd metric is that it is not possible to apply this methodology to 2D queries. It also has the disadvantage that the geometric value of a fragment is no longer an intrinsic property of the fragment, but rather applies only in relation to a specific query fragment. In other words, the pairwise torsion rmsd metric does not give us any information on how the hits relate to each other. This is not just because the value relates to a specific query, but also because we have reduced the multidimensional conformational description to a single scalar value. For example, given two hit fragments, each with an rmsd of 30° to the query, we do not know whether these two fragments are identical to each other or are 60° apart from each other (or more likely, somewhere in between). In order to overcome this, Mogul has built-in functionality for superimposing and visualizing the hit rings. This functionality is useful for understanding the conformational preferences of the ring system under investigation. However, if there are a large number of rings, then it can be difficult to identify distinct ring conformations. An option to cluster the rings based on their rmsd to each other has therefore been implemented. A standard hierarchical agglomerative clustering procedure is applied, using the torsion rmsd measure described above as the pairwise distance measure.

Independent CSD Data Set. An independent test set of CSD structures not included in the Mogul knowledge base was created by searching for five- and six-membered rings in the November 2010 CSD data update. Two ConQuest searches, one for five-membered and one for six-membered rings, were constructed. These were set up to allow any bond types between the atoms. The atom types were restricted to exclude metal atoms. Furthermore, restraints were imposed to exclude rings forming part of fused ring systems. The searches also applied the following filters: $R \leq 0.1$, "3D coordinates determined", "not disordered", "no errors", "not polymeric", "no ions", "no powder structures", "only organics". In order to exclude aromatic rings, the sum of the absolute torsions in each retrieved ring was calculated using the data analysis functionality within the program Mercury,[27] and the ring was discarded if this value was less than 15° for five-membered rings or 18° for six-membered rings.

Settings Used in the Examples. All Mogul searches were carried out using the "exclude organometallics" filter.

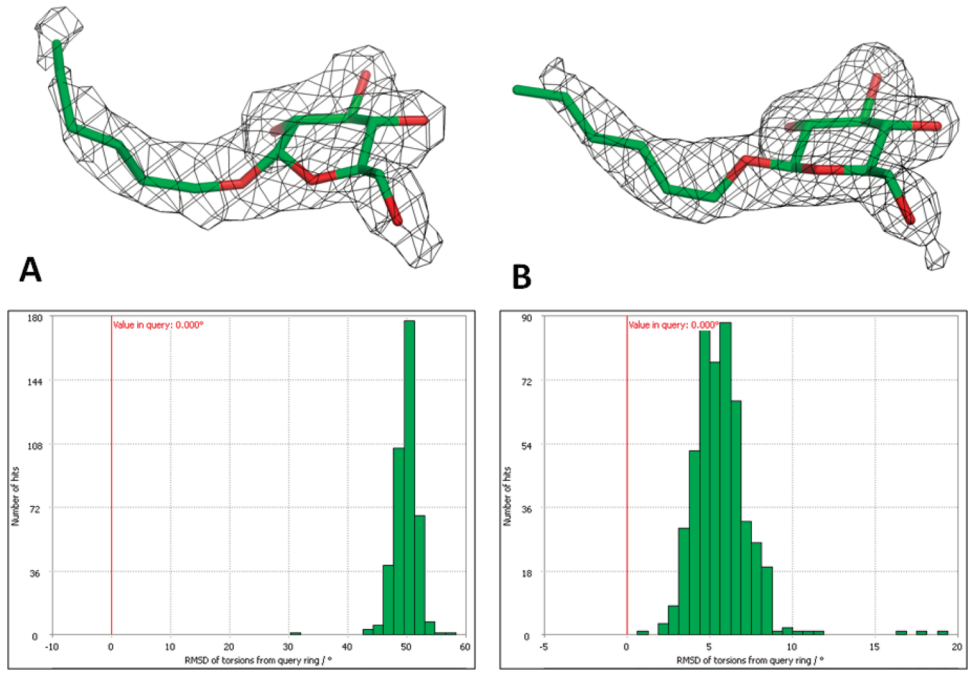


Figure 2. Original (left) and re-refined (right) structures of the ligand (PDB code 2evs). In the re-refined structure, the glucoside ring takes on a chair conformation. Note the improved torsion rmsd metric for the re-refined structure.

RESULTS AND DISCUSSION

Identifying Unusual Ring Geometries. We envisage that the primary use of the ring functionality in Mogul will be to validate ring geometries of ligands modeled using computational methods or fitted to electron density in the determination of a protein−ligand complex structure. Using the torsion rmsd metric, it is possible to classify rings as usual or unusual. Ring conformations are classified as unusual if less than 5% of the experimental values are within 10° of the conformation in the query.

As a first test, we looked at an independent test set of CSD structures not forming part of the Mogul knowledge base. The expectation was that a minority of these structures would be highlighted as unusual, as compounds in small molecule structures tend to adopt low energy conformations.[21] The November 2010 CSD update contained 358 (150 five-membered and 208 six-membered) structures with nonflat rings. Analysis of these structures revealed that only one of them had an unusual ring (KUPNEP[28]). The KUPNEP crystal structure consists of *p-tert*-butyl-calix[8]arene crystallized with six *N*-methyl-morpholine solvent molecules. Three of the six *N*-methyl-morpholine rings are highlighted as unusual. It would appear that the authors have refined the structure without the use of restraints, and inspection of the structure suggests that the original data may not have been of high enough quality to resolve the *N*-methyl-morpholine atomic positions freely.

Let us now illustrate the use of the ring functionality in Mogul to validate the geometry of a ligand structure in a protein. The asymmetric unit of the protein structure human placental annexin V (PDB[29] code 1hak[30]), contains two protein chains each binding a piperidine containing ligand. The conformation of the piperidine ring of the ligand binding to chain A is modeled differently to that binding to chain B (Figure 1). The chain A piperidine ring is in a chair conformation whereas that of chain B is in a twist-boat conformation. Performing a search on the chair conformation reveals that the conformation is not unusual. The rmsds with respect to the data in the CSD lie in the range of 2−6°. The twist-boat conformation, on the other hand, is classified as unusual with rmsds in the range of 69−74° (no experimental data falls within 10° of this conformation (Figure 1)).

Let us now look at the protein−ligand complex of human glycolipid transfer protein with *N*-hexyl-β-D-glucoside, which contains a glucoside ring in an envelope conformation (PDB code 2evs[31]). Mogul reports the envelope ring conformation as unusual, the torsion rmsds lie in the range of 31−56°, but inspection of the ligand in the electron density, using the Uppsala Electron Density Server,[32] reveals a relatively good fit of the ligand (Figure 2A). However, closer examination reveals that the ligand modeled into the 2evs structure is the α-anomer, whereas the original study was carried out on the β-anomer. Rerefining the 2evs structure with the β-anomer using BUSTER[33] and grade[34] gave an alternative model that fits the electron density equally well (Figure 2B). In the rerefined structure, the glucoside ring adopts a chair conformation. Furthermore, using the rerefined structure Mogul reports the ring as nonexceptional (the torsion rmsds lie in the range of 1−19°).

It is important to note that Mogul identifies only whether ring conformations are usual or unusual, not whether they are correct or incorrect. It is up to the user to make the latter decision. One example of an unusual ring conformation that is believed to be correct is the tetrahydropyran ring of CSD entry LAFFAZ.[35] In this case the tetrahydropyran ring adopts a twist-boat conformation (Figure 3) and the rmsds to the other ring conformations range from 45−68°. A similar conformation of the ring has been observed when bound to glycogen phosphorylase.[36] Both anomeric and steric effects have been postulated to contribute to this unusual conformation.[35] It is interesting to note that in this particular case the small molecule crystal structure was solved in order to be able to

dx.doi.org/10.1021/ci200439d | J. Chem. Inf. Model. 2012, 52, 956−962

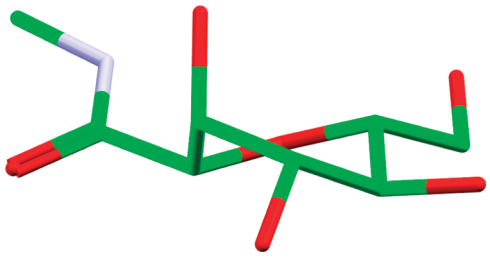


Figure 3. Twist-boat conformation of the tetrahydopyran ring (CSD refcode LAFFAZ).

provide further evidence for the unusual conformation observed in the glycogen phosphorylase structure.

Another interesting example of an unusual ring geometry is the central *N*-acetyl-D-glucosamine (NAG) residue of chitin bound to the catalytic domain of chitinase (PDB code 3a4w[37]). In this structure the NAG ring is distorted into a 1,4B boat conformation (Figure 4), which is highlighted as unusual (rmsds ranging from 46° to 75°). In this example the boat conformation adopted by the central NAG residue is correct; in fact it is a consequence of the catalytic mechanism of the enzyme which creates a high energy state for the ligand which then becomes much more susceptible to hydrolysis.[38]

Search Generalization. The CSD hits retrieved by a Mogul search are, by design, very specific to the chemical environments that surround the geometric features being probed. This can result in few or no hits if a particular chemistry is not well represented in the CSD. In these cases one can generalize the search for chemistry of interest. Take for example the structure of loracarbef bound to the AmpC beta-lactamase mutant Q120L/Y150E (PDB code 1fcn[39]). When covalently bound to the beta-lactamase, the fused ring of loracarbef is split and a tetrahydropyridine ring is revealed. Using Mogul to look at look at the conformation of this ring, we do not find any rings that have an identical substitution pattern. However, we are able to find hits with the same atom types in the ring. In this example six entries from the CSD have three ring atoms with the same substitution pattern as the query and nine CSD entries where two ring atoms have the same substitution pattern (Figure 5). As the torsion rmsds of these hits fall in the range of 3−23° we conclude this ring has reasonable geometry.

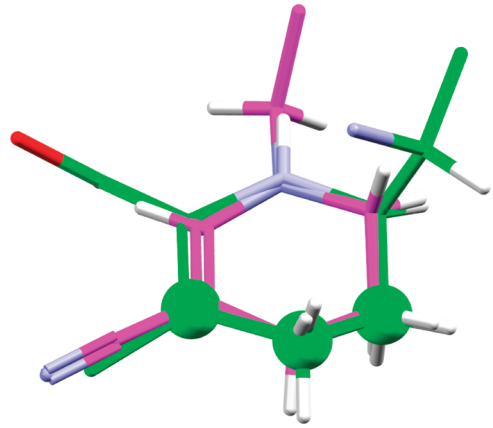


Figure 5. Tetrahydropyridine ring of CSD refcode ZUHSIE[40] (magenta) superimposed on Loracarbef (green) using the heavy atoms of the ring. In this instance three of the ring atoms (green spheres) have the same substitution pattern.

Clustering of Hits. The conformational flexibility of a particular ring system can be evaluated by superimposing and clustering the ring structures retrieved from the CSD. Take for example the dihydropyran ring of the neuraminidase inhibitor zanamivir (PDB code 3b7e[41]). In this case the ring conformation reported in the protein structure is not unusual, the torsion rmsds of the query with respect to the rings retrieved from the CSD lie in the range of 4−15°. Splitting the data into two clusters reveals that there is not much variation in the ring structure. The torsional rmsd distance between the two clusters is only 13.8°. Visual inspection reveals that the query ring is more similar to the first cluster (represented by GUVFAE in Figure 6) than the second cluster (represented by VOFJOQ in Figure 6). Furthermore, visual inspection of the superimposed ring clusters reveals that, although this is a fairly rigid ring structure, there is some degree of flexibility in the positioning of the equatorial C1 substituent (Figure 6).

Let us now look at cycloheptane with a single primary substituent. In this case 14 hits are retrieved, and 3 main

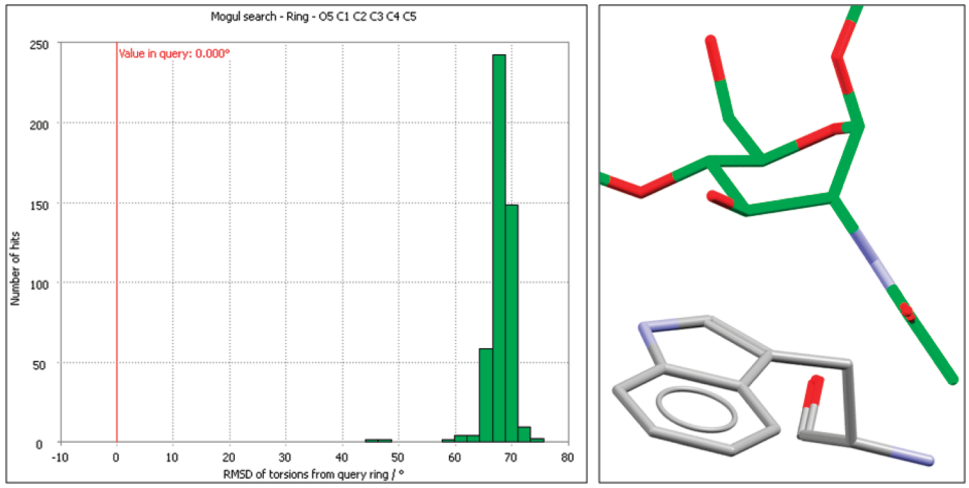


Figure 4. Central NAG residue of a NAG pentamer adopts a boat conformation which is stabilized by hydrophobic stacking interactions to TRP664 in the chitinase binding site (PDB code 3a4w). Mutating this tryptophan to alanine reduces the k_{cat}/K_m value by almost 4 orders of magnitude.[37]

dx.doi.org/10.1021/ci200439d | J. Chem. Inf. Model. 2012, 52, 956−962

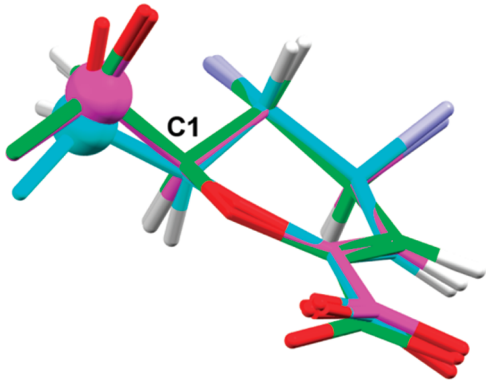


Figure 6. Dihydropyran rings of CSD refcodes GUVFAE (magenta) and VOFJOQ (cyan) superimposed on the zanamivir ring (PDB code 3b7e, green). Note the degree of flexibility in the positioning of the equatorial C1 substituent (sphere).

clusters are identified. The largest of the clusters contains conformations ranging from the chair to the twist-boat conformation. In these conformations, we see how the absolute value of torsion 7, using the numbering scheme of Allen et al.,[10] ranges from 8° to 39°, where 0° represents an ideal chair conformation and 37.3° represents an ideal twist-boat conformation. The other two clusters each contain a single entry and represent a boat (TEQOM) and an envelopelike (VOBNEF) conformation, respectively. Two main conclusions can be drawn from these results. First, cycloheptane with a single primary substituent prefers to adopt chair/twist-chair conformations. Second, a spectrum of conformations between the chair and the twist-chair conformations is observed (Figure 7).

CONCLUSION

We have described the creation of a ring geometry knowledge base from the CSD. This knowledge base, accessible via the Mogul module of the CSDS, allows rapid access to geometric information on rings. The ring tree library contains over 900 000 fragments. The treelike structure of the data makes fragment searches extremely fast. Furthermore, it makes it possible to generalize searches in cases where no exact hits are found. In order to reduce the inherent complexity of comparing ring structures, a single scalar value, the torsion rmsd, is used to compare the query ring conformation to the CSD ring fragment hits. Superimposing and clustering of hit ring fragments is also available in Mogul. This allows the user to get an immediate understanding of the conformational preferences of the fragment ring under investigation.

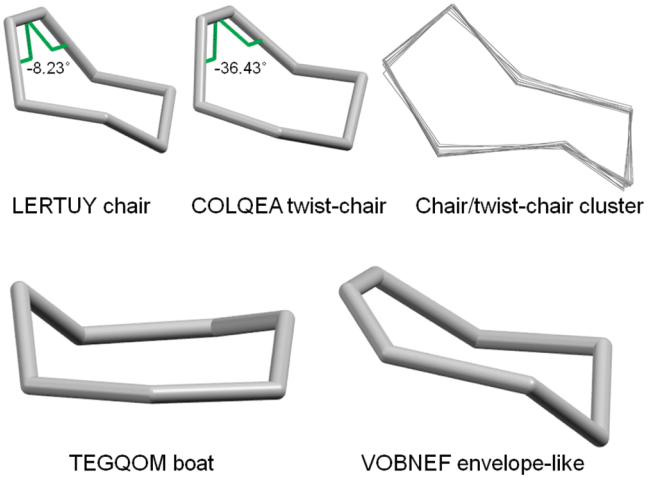


Figure 7. Three main clusters are identified for cycloheptane with a single primary substituent. The top row shows the chair/twist-chair cluster and two representative conformations. Note that the absolute value of torsion 7 (highlighted) takes on a range of values between 8° and 39°. The bottom row shows the structures from the other two clusters.

Nonaromatic rings provide useful scaffolds for flexibly directing functionality into different areas of a protein active site. Although synthesis of such molecules can often be challenging it could prove to be a worthwhile endeavor. Studies have shown that increasing the aromatic ring count can have detrimental effects on compound developability[42] and that clinically approved drugs have a more saturated character than compounds in earlier stages of development.[43] Accurate modeling and validation of the conformations of nonaromatic rings presents a strategy for gaining confidence in the predictions made using compounds containing them. The use of knowledge from the CSD to validate any modeled ring geometries is therefore of high value in this regard. Fused and bridged ring systems form an important part of many druglike molecules. These have not so far been considered in our work, as the key definitions and torsion rmsd metric described here cannot be applied to them in their current form. However, a natural extension of the work presented here would be the development of a method to analyze fused and bridged ring systems. Another useful development would be the ability to compare rings that have differing atom types but are electronically similar.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cottrell@ccdc.cam.ac.uk.

Present Address

[†]Taylor Cheminformatics Software, 54 Sherfield Avenue, Rickmansworth, Herts, WD3 1NL, U.K.

ACKNOWLEDGMENTS

We thank Oliver Smart, Thomas Womack, and Andrew Sharff at Global Phasing for rerefining the 2evs structure, highlighting the 3a4w example and critical reading of the manuscript. We also thank Colin Groom for critical reading of the manuscript and the reviewers for their constructive feedback.

REFERENCES

(1) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.

(2) Xu, J. A new approach to finding natural chemical structure classes. *J. Med. Chem.* **2002**, *45*, 5311−5320.

(3) Feher, M.; Schmidt, J. M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218−227.

(4) Lee, M. L.; Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **2001**, *3*, 284−289.

(5) Lipkus, A. H. Exploring chemical rings in a simple topological-descriptor space. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 430−438.

(6) Nilakantan, R.; Bauman, N.; Haraki, K. S. Database diversity assessment: New ideas, concepts, and tools. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 447−452.

(7) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; Mclay, L. M.; Bradshaw, J. Drug rings database with web interface. A tool for identifying alternative chemical rings in lead discovery programs. *J. Med. Chem.* **2003**, *46*, 3257−3274.
(8) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic rings of the future. *J. Med. Chem.* **2009**, *52*, 2952−2963.
(9) Brameld, K. A.; Kuhn, B.; Reuter, D. C.; Stahl, M. Small molecule conformational preferences derived from crystal structure data. A medicinal chemistry focused analysis. *J. Chem. Inf. Model.* **2008**, *48*, 1−24.
(10) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. 1. Cycloheptane. *Acta Crystallogr., Sect. B* **1993**, *49*, 910−928.
(11) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A.; Vinter, J. G. Symmetry-modified conformational mapping and classification of the medium rings from crystallographic data. 2. Exo-unsaturated and heterocyclic 7-membered rings. *Acta Crystallogr., Sect. B* **1994**, *50*, 382−395.
(12) 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. 3. Endo-unsaturated 7-membered rings. *Acta Crystallogr., Sect. B* **1994**, *50*, 395−404.
(13) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. The thermodynamics and molecular structure of cyclopentane. *J. Am. Chem. Soc.* **1947**, *69*, 2483−2488.
(14) Sadowski, J.; Gasteiger, J. From atoms and bonds to 3-dimensional atomic coordinates − Automatic model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(15) Gerber, P. R.; Gubernator, K.; Muller, K. Generic shapes for the conformation analysis of macrocyclic structures. *Helv. Chim. Acta* **1988**, *71*, 1429−1441.
(16) Goto, H.; Osawa, E. An efficient algorithm for searching low-energy conformers of cyclic and acyclic molecules. *J. Chem. Soc., Perkin Trans. 2* **1993**, 187−198.
(17) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.
(18) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of automatic 3-dimensional model builders Using 639 X-ray structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(19) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534−546.
(20) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.
(21) Allen, F. H.; Harris, S. E.; Taylor, R. Comparison of conformer distributions in the crystalline state with conformational energies calculated by ab initio techniques. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 247−254.
(22) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr., Sect. A* **1991**, *47*, 392−400.
(23) Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; Orpen, A. G. Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2133−2144.
(24) Cremer, D.; Pople, J. A. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.
(25) Tripos Mol2 File Format. http://tripos.com/data/support/mol2.pdf (accessed December 19, 2011).
(26) Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. Errors in protein structures. *Nature* **1996**, *381*, 272−272.
(27) Sykes, R. A.; McCabe, P.; Allen, F. H.; Battle, G. M.; Bruno, I. J.; Wood, P. A. New software for statistical analysis of Cambridge Structural Database data. *J. Appl. Crystallogr.* **2011**, *44*, 882−886.
(28) Mecca, T.; Consoli, G. M. L.; Geraci, C.; Camalli, M.; Campi, G.; Spagna, R. Crystal structure of a p-tert-butylcalix[8]arene - N-methyl-morpholine complex. *Z. Kristallogr.—New Cryst. Struct.* **2009**, *224*, 407−411.
(29) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. Protein Data Bank − Computer-based archival file for macro-molecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(30) Kaneko, N.; Ago, H.; Matsuda, R.; Inagaki, E.; Miyano, M. Crystal structure of annexin V with its ligand K-201 as a calcium channel activity inhibitor. *J. Mol. Biol.* **1997**, *274*, 16−20.
(31) Malinina, L.; Malakhova, M. L.; Kanack, A. T.; Lu, M.; Abagyan, R.; Brown, R. E.; Patel, D. J. The liganding of glycolipid transfer protein is controlled by glycolipid acyl structure. *PLoS Biol.* **2006**, *4*, 1996−2011.
(32) Kleywegt, G. J.; Harris, M. R.; Zou, J. Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The Uppsala Electron-Density Server. *Acta Crystallogr., Sect. D* **2004**, *60*, 2240−2249.
(33) Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O. S.; Vonrhein, C.; Womack, T. O. *BUSTER*, version 2.9; Global Phasing Ltd.: Cambridge, U.K., 2010.
(34) Smart, O. S.; Womack, T. O.; Sharff, A.; Flensburg, C.; Keller, P.; Paciorek, W.; Vonrhein, C.; Bricogne, G. *Grade*; Global Phasing Ltd.: Cambridge, U.K., 2011.
(35) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J. F.; Fleet, G. W. J.; Ford, P.; Watkin, D. J.; Oikonomakos, N. G. X-Ray crystallographic analysis of 2,6-anhydro-N-methyl-D-glycero-D-ido-heptonamide - the first example of a simple glucose analog with a skew boat structure. *J. Chem. Soc., Chem. Commun.* **1993**, 654−656.
(36) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J. F.; Orchard, M. G.; Fleet, G. W. J.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M.; Papageorgioui, A. Design of inhibitors of glycogen-phosphorylase - a study of alpha-C-glucosides and beta-C-glucosides and 1-thio-beta-D-glucose compounds. *Biochemistry* **1994**, *33*, 5745−5758.
(37) Tsuji, H.; Nishimura, S.; Inui, T.; Kado, Y.; Ishikawa, K.; Nakamura, T.; Uegaki, K. Kinetic and crystallographic analyses of the catalytic domain of chitinase from Pyrococcus furiosus - the role of conserved residues in the active site. *FEBS J.* **2010**, *277*, 2683−2695.
(38) van Aalten, D. M. F.; Komander, D.; Synstad, B.; Gaseidnes, S.; Peter, M. G.; Eijsink, V. G. H. Structural insights into the catalytic mechanism of a family 18 exo-chitinase. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8979−8984.
(39) Patera, A.; Blaszczak, L. C.; Shoichet, B. K. Crystal structures of substrate and inhibitor complexes with AmpC beta-lactamase: Possible implications for substrate-assisted catalysis. *J. Am. Chem. Soc.* **2000**, *122*, 10504−10512.
(40) Hussain, Z.; Fleming, F. F.; Norman, R. E.; Chang, S. C. 3-cyano-1-[4-(1,3-dithian-2-yl)butyl]-1,4,5,6-tetrahydropyridine. *Acta Crystallogr., Sect. C Commun.* **1996**, *52*, 1010−1012.
(41) Xu, X. J.; Zhu, X. Y.; Dwek, R. A.; Stevens, J.; Wilson, I. A. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J. Virol.* **2008**, *82*, 10493−10501.
(42) Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discovery Today* **2011**, *16*, 164−171.
(43) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, *52*, 6752−6756.