

PERSPECTIVE

Small Molecule Conformational Preferences Derived from Crystal Structure Data. A Medicinal Chemistry Focused Analysis

Ken A. Brameld,[†] Bernd Kuhn,[‡] Deborah C. Reuter,[†] and Martin Stahl^{*,‡}

Discovery Chemistry, Roche Palo Alto LLC, Palo Alto, California, and Discovery Chemistry, F. Hoffmann-La Roche AG, Basel CH-4070, Switzerland

Based on torsion angle distributions of frequently occurring substructures, conformation preferences of druglike molecules are presented, accompanied by a review of the relevant literature. First, the relevance of the Cambridge Structural Database (CSD) for drug design is demonstrated by comparing substructures present in compounds entering clinical trials with those found in the CSD and protein-bound ligands in the Protein Data Bank (PDB). Next, we briefly highlight preferred conformations of elementary acyclic systems, followed by a discussion of sulfonamide conformations. Due to their central role in medicinal chemistry, we discuss properties of aryl ring substituents in depth, including biaryl systems and systems of two aryl rings connected by two acyclic bonds. For a subset of torsion motifs, we also compare torsion angle histograms derived from CSD structures with those derived from ligands in the PDB. Furthermore, selected properties of some six- and seven-membered ring systems are discussed. The article closes with a section on attractive sulfur–oxygen contacts.

1. INTRODUCTION

The ensemble of conformations a small molecule can access is critical to many aspects of drug discovery, since these conformations define the shape of the molecule, which in turn is a major determinant of its biological and physical properties. Alternative conformations of organic molecules generally differ in the torsion angles of their rotatable single bonds. Already few rotameric positions per bond and few alternate ring geometries lead to a large number of theoretically possible conformations that a low molecular weight molecule could in principle attain. Most of these conformations are energetically unfavorable and are not observed experimentally. The challenge faced by a medicinal chemist is to understand the structural and energetic differences between these conformations so as to successfully design new molecules that adopt the desired shape with little or no energetic penalty. Energy differences between conformational states are related thermodynamically to their populations, with a ratio of 1:10 for a free energy difference of ≈ 1.4 kcal/mol at room temperature. Consequently, conformations with relative free energies larger than 2 kcal/mol are populated to a very minor extent. This review article addresses the use of small molecule crystal structures for drug design with a specific focus on the observed conformations for common medicinal chemistry organic fragments.

Conformational analysis of small molecules is a critical component for all 3D computer-aided drug design tools and should be of concern to any medicinal chemist attempting

to do drug design. Computational conformational analysis is a central task in molecular design (for a review see ref 1).

Experimentally, small molecule conformations are commonly studied using NMR and X-ray crystallography. The Cambridge Structural Database (CSD)² is the primary repository for small molecule crystallographic data and the principal focus of this article. It currently contains over 390 000 structures determined by X-ray or neutron diffraction and continues to grow rapidly. For considerable time, the CSD has served as a key data source on molecular interaction patterns, reaction pathways, conformations, and for scaffold hopping in drug design.^{3–8} An overview of the relevant literature can be obtained through the CCDC Web page.⁹ The value of the CSD for conformational analysis stems from the large number of observations for different organic fragments in different crystal packing environments. By examining the experimentally observed torsion angle distribution for an organic fragment of interest, it is possible to determine its preferred geometry. As an alternative to analyzing small molecule crystal structures, one may examine the conformations of ligands bound to their protein targets by studying the macromolecular crystal structures deposited in the Protein Data Bank (PDB).^{10,11} These data are generally of lower resolution and the ligands cover less chemical diversity. However, both sources of ligand conformational geometries are addressed in this work.

This article is divided into two parts that address the conformational preferences observed in the CSD and PDB for pharmaceutically relevant organic fragments. The first part examines the relevance of studying the CSD conformations of a ligand as a means to gain insights into the solution phase or target-bound structure. Furthermore, the coverage of pharmaceutical chemical space is assessed by comparing

* Corresponding author phone +41 61 68 88421; e-mail: martin.stahl@roche.com. Corresponding author address: F. Hoffmann-La Roche AG, PRBD-CM 92/3.10B, CH-4070 Basel, Switzerland.

[†] Discovery Chemistry, Roche Palo Alto.

[‡] Discovery Chemistry, Roche Basel.

the molecular cores found in compounds that have entered clinical trials to those contained in the CSD and in ligand–protein complexes deposited in the PDB. The second part, organized in several thematic subsections, reviews the torsion angle distributions for organic fragments commonly used in medicinal chemistry. This survey is not exhaustive but instead covers a representative set of organic fragments routinely accessed by medicinal chemists and should be of interest to anyone engaged in the challenging endeavor of molecular design.

2. RELEVANCE OF SMALL MOLECULE CRYSTAL STRUCTURE CONFORMATIONS FOR DRUG DESIGN

In the context of drug design, the conformation a small molecule adopts when bound to a pharmaceutical target is of fundamental importance. One may wonder how relevant the conformations observed in small molecule crystals are to the bioactive conformation, as energy differences between alternative conformations can be overcome by the formation of favorable nonbonded interactions. A particular small molecule conformation observed in a crystal may thus be due to an unusually favorable intermolecular interaction that is not present in solution. The risk of being misled by such circumstances can be minimized by drawing conclusions only from searches that yield a statistically significant number of observations, since any conformational strain reduces the total crystal free energy and its occurrence is therefore not a likely event. However, systematic differences may occur between the solid phase and solution or vapor phases. The conformation of biphenyl is the classic example of crystal-field effects leading to a preference for a coplanar geometry, whereas the vapor-phase structure has a torsion angle of 44° .¹² Despite this example, the validity of using conformer distributions in the crystalline state is now fairly well established. Studies comparing the conformational energies calculated by *ab initio* methods to observed torsion distributions routinely show a good correlation. A systematic study of 12 organic substructures found that torsion angles with high strain energy (>1 kcal/mol) are rarely observed in crystal structures.¹³ Indeed, one conclusion from that study is that “high-energy conformers are underrepresented in crystal structures compared with gas-phase, room-temperature Boltzmann distributions”. Consequently, one may err on the side of being too conservative in considering only crystallographic distributions when defining the preferred conformations for organic fragments.

Several studies indicate that small molecule crystal structure conformations correlate well with protein-bound conformations. Boehm and Klebe comment that for polar ligands in particular, a single crystal structure of a ligand alone is not a reliable predictor of the bioactive conformation.¹⁴ Rather than relying on a single structure, more robust results can be attained when considering multiple structures which sample a variety of local environments. With the exception of common buffer components, it is rare to have a large number of structures for the same molecule in different environments. It is a more common situation to observe the same molecular substructure in many crystal structures of different compounds, so more general conclusions might be drawn from the study of molecular fragments. One such study of 80 protein–ligand complexes revealed

similar torsion angle distributions for fragments when the bound and unbound distributions were compared.^{14,15}

Several conformer generation programs employ empirically derived torsion libraries to help ensure that only energetically favorable conformations are produced. The transferability of CSD torsion angle distributions to biologically active ligand conformations is borne out in the success of these programs. A recent study using the program OMEGA¹⁶ allowed for a direct comparison of the benefits of including empirical data from the CSD.¹⁷ OMEGA uses a torsion library to define the allowable torsion angle values for 4-atom fragments. This default library has specific rules for 123 SMARTS¹⁸ patterns, of varying specificity. An automated and systematic study of organic molecules in the CSD was used to generate a much more specific torsion library with 1864 SMARTS patterns. For a test set of approximately 8500 CSD crystal structures not used in the development of the new torsion library, the default torsion library predicted a conformer within 0.5 \AA root-mean-square deviation of the crystal structure in 84% of the cases. The new torsion library improved this success rate to 92%. A comparison for a set of 1267 protein-bound ligands showed a similar improvement in success rate from 70% to 79%. For both CSD and PDB test sets, the new CSD-derived torsion library led to conformers that were closer to the experimental structures.

These studies provide a compelling argument in favor of using the structural data contained in the CSD for drug design applications. In general, the conformations observed in the CSD are of low energy when compared to *ab initio* molecular orbital calculations. These small molecule crystal conformations are comparable to the protein-bound conformations and therefore are directly relevant to structure-based drug design. As the number of structures in the CSD has continued to grow, it is now reasonable to assess if the database is of sufficient size and diversity to address many of the organic fragments frequently employed by medicinal chemists.

3. MATERIALS AND METHODS

All statistical data on small molecule crystal structures described in this article was derived by searching CSD Version 5.28 (November 2006) with the ConQuest 1.9 program.¹⁹ Unless noted otherwise, the following general search flags were set: R factor ≤ 0.10 , “3D coordinates determined”, “not disordered”, “no ions”, “no errors”, “not polymeric”, and “only organic”. This results in a search space of 103 896 structures. Histograms for individual torsion angles were obtained by exporting query results to Excel²⁰ and running a macro using the “frequency” array function. All histograms are generated such that each bin of 5° counts all structures with torsion angles a in a range $x < a \leq x + 5^\circ$, where x is the lower bound of the bin. For $x = 0^\circ$, all structures with values of exactly 0° are also included, i.e., the first bin counts all structures with $0^\circ \leq a \leq 5^\circ$. Torsion angle searches with ConQuest return values between -180° and $+180^\circ$. Frequency histograms between -180° and 0° can be regarded as equivalent to their mirror images with corresponding positive torsion angle values. Throughout the paper, we have thus “folded” the negative part of the distribution onto the positive one by plotting absolute torsion angle values. This leads to histograms in the range of 0°

and 180° which are in general statistically more meaningful. In the case of symmetric substructures with redundant torsion angle definitions (e.g., substituents on aryl rings where no *ortho* substituents are present), one arbitrary torsion angle was chosen, and the same procedure was applied. Quantum mechanical calculations referred to in the text were performed with Gaussian 98²¹ except for the hypersurfaces depicted in Figure 18, which were generated with VAMP.²²

Conformations of protein-bound ligands were analyzed by extracting ligands in SD format²³ from Proasis2,²⁴ a curated version of the PDB,^{10,11} and generating a separate database in CSD format with the program PreQuest.²⁵ The ligand structures were taken from all HET entries, except metals and commonly found small ions, in the PDB as of June 1, 2007. Those with less than 5 or more than 100 atoms were removed, excluding in particular large peptidic ligands. Multiple occurrences of the same ligand in one PDB structure were counted as separate entries. Thus, the database contains a total of 62 010 ligands derived from 19 664 PDB entries. Only 5.1% of structures have resolutions above 3.0 Å. 59.4% of the database entries have resolutions of 2.25 Å and better. Queries in this PDB database were run twice, once using all structures and once excluding structures with a resolution above 2.25 Å (the choice of this limit is due to technical constraints of the ConQuest system, which allows the use of certain fixed resolution ranges only). We generally see no difference between the frequency distributions resulting from these two types of queries, which is not surprising given that “resolution” is primarily a measure of data quantity and not of coordinate precision.²⁶ A higher precision might generally be expected from a resolution of below 1.7 Å, but for most substructures this criterion dramatically reduces the number of observations. In the article we therefore report query results for the full database only. Queries were run in an identical manner in the CSD and the PDB ligand database, with the exception that for CSD searches explicit hydrogen definitions were used, whereas the absence of hydrogen atoms in the PDB ligands required the use of implicit hydrogen counts.

For the molecular cores analysis, a subset of compounds contained in CSD Version 5.28 was obtained by applying filters similar to those described by Sadowski and Bosstroem.¹⁷ Briefly, the flags available in the ConQuest interface were set as described above. Additional filters were applied to remove salts and restrict the set to structures with a molecular weight range between 100 and 800, containing only the elements H, C, N, O, F, S, P, Cl, Br, and I. A maximum of 10 rotatable bonds and a ring size of up to 9 atoms were permitted. This resulted in a final set of 79 744 molecules with crystallographic coordinates which will be referred to as the restricted CSD. A similar analysis of the PDB ligand database resulted in 39 850 ligands and is referred to as the restricted PDB.

The Prous Integrity database²⁷ was used as a source of medicinal chemistry compounds that have progressed into Phase I clinical trials or beyond, including currently marketed drugs. The following keywords were used as search criteria: “Phase I” or “Phase I/II” or “Phase II” or “Phase II/III” or “Phase III” or “Pre-Registered” or “Recommended Approval” or “Registered” or “Launched” and molecular weight 250–800. This search returned 4286 hits. Similar filters to those applied to the CSD were used to remove salts and any

entries with elements other than H, C, N, O, F, S, P, Cl, Br, and I. This resulted in a final set of 3988 unique molecules.

The molecular cores present in the restricted CSD or PDB entries and the Prous structures were identified using software developed internally with the OEchem toolkit.¹⁶ Molecules were dissected into ring systems, side-chain atoms, and linker atoms as defined by Bemis and Murcko²⁸ with some minor modifications. The primary difference in this work is the inclusion of exocyclic double-bonded terminal atoms in the definition of ring systems or linker atoms. As a consequence, amide and ester linkers include the carbonyl oxygen in the linker definition. In this way, a thioamide is identified as a different linker than an amide. Similarly, a cytosine ring system will include the exocyclic carbonyl as part of the ring system. In the current work, a molecular core is defined to be a minimum of a ring-linker-ring system, while retaining the element and bond orders as found in the original neutral molecule. For our purpose of studying conformational flexibility, molecular cores were retained which had a maximum molecular weight of 300 and a maximum of 4 rotatable bonds.

4. CRYSTALLOGRAPHIC COVERAGE OF PHARMACEUTICAL CHEMICAL SPACE

The CSD is a collection of small molecule crystal structures from a diverse set of sources and with no specific focus toward pharmaceuticals. Therefore, it is reasonable to probe the relevance of its content to the field of medicinal chemistry. While many analyses have been reported which attempt to identify the characteristics that distinguish drugs from nondrugs,^{29,30} for our intended analysis of conformational preferences, we are not strictly interested in identifying the number of druglike molecules in the CSD or PDB. It is of greater importance to understand the coverage of conformationally flexible organic fragments commonly used in medicinal chemistry (for a related analysis in the context of scaffold hopping see ref 5). Organic fragments that serve as linkers between ring systems are of particular interest as they have the greatest potential to impact the shape of a molecule. The identity and coverage of these ring-linker-ring systems were studied in greater detail.

The Prous Integrity database²⁷ was used to identify a set of compounds that have progressed into Phase I clinical trials or beyond and therefore are of importance to medicinal chemists. This set of 3988 unique compounds was then broken down into molecular cores which were required to be two or more ring systems, optionally linked by one or more acyclic bonds. The final molecular core was restricted to contain a maximum of four rotatable bonds and a molecular weight less than 300. The algorithm used to identify the molecular cores retains the original bond order and elements of the parent molecule; thus the local environment of the linker between the rings is maintained. A total of 2308 unique molecular cores were identified, and the frequency of each was recorded. This set of cores will be referred to as drug molecular cores. The 20 most frequent drug molecular cores are shown in Figure 1. The identity and frequency of these cores is in qualitative agreement with a similar analysis of molecular frameworks found in known drugs.²⁸

We are now in a position to ask how many of these drug molecular cores have structural representation in the CSD

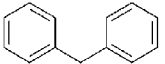
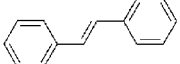
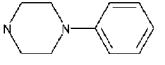
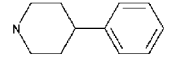
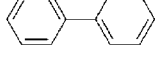
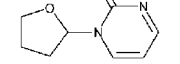
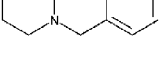
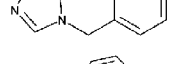
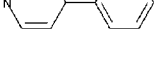
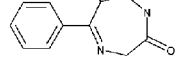
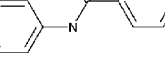
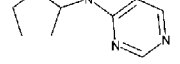
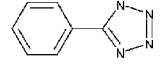
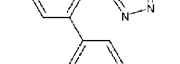
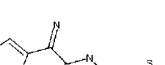
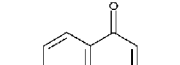
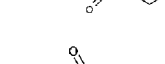
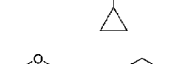
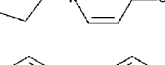
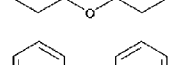
Structure	Frequency			Structure	Frequency		
	Prous	CSD	PDB		Prous	CSD	PDB
	110 (2.50%)	2250 (2.60%)	43 (0.20%)		24 (0.55%)	560 (0.65%)	97 (0.44%)
	69 (1.57%)	88 (0.10%)	27 (0.12%)		24 (0.55%)	38 (0.04%)	18 (0.08%)
	67 (1.52%)	1213 (1.40%)	135 (0.62%)		23 (0.52%)	80 (0.09%)	391 (1.79%)
	60 (1.36%)	61 (0.07%)	16 (0.07%)		23 (0.52%)	82 (0.09%)	19 (0.09%)
	43 (0.98%)	108 (0.12%)	0 (0.00%)		23 (0.52%)	47 (0.05%)	2 (0.01%)
	31 (0.71%)	238 (0.28%)	34 (0.16%)		23 (0.52%)	80 (0.09%)	5353 (24.5%)
	29 (0.66%)	25 (0.03%)	1 (0.00%)		22 (0.50%)	10 (0.01%)	1 (0.00%)
	29 (0.66%)	1 (0.00%)	0 (0.00%)		20 (0.45%)	15 (0.02%)	3 (0.01%)
	26 (0.59%)	176 (0.20%)	1174 (5.37%)		20 (0.45%)	56 (0.06%)	3864 (17.7%)
	24 (0.55%)	269 (0.31%)	150 (0.69%)		19 (0.43%)	271 (0.31%)	29 (0.13%)

Figure 1. The 20 most frequently observed molecular cores for compounds entering Phase I clinical trials or beyond, as identified in the Prous Integrity database. The absolute observation frequencies are reported for each database (Prous, CSD, PDB) followed by the relative frequencies in parentheses. Note the considerable relative frequency of nucleosides and sugars in the PDB (numbers in bold print).

or PDB. Given the diverse nature of the compounds contained in the CSD, some care was taken to exclude entries that may contain conformational artifacts. Therefore a restricted subset of the CSD was defined primarily to exclude metal-containing structures, macrocyclic systems, and extremely large molecules, as defined in the Methods section. From the remaining 68 016 CSD molecules, 29 179 unique ring-linker-ring molecular cores were identified. A similar analysis of ligands extracted from Proasis2,²⁴ a curated version of the PDB, yielded only 1726 unique molecular cores from 39 850 ligands. This relatively small number of PDB molecular cores is a striking manifestation of the lack of chemical diversity for ligand–protein complexes in the PDB. Indeed, the 20 most frequent PDB molecular cores cover 70% of the ligand–protein complexes and are dominated by nucleosides, nucleotides, and sugars.

Of the 2308 unique drug molecular cores from the Prous compounds, 845 identical matches are present in the CSD. At first glance, this 37% coverage may seem somewhat low. However, when the observation frequency for each molecular

core is considered, it becomes clear that the coverage is poor primarily due to rare drug molecular cores, present in only a single Prous entry (i.e., singletons). Figure 2 plots the CSD and PDB coverage as a function of observation frequency in drugs. This plot clearly shows that the molecular cores that occur with low frequency in drugs also have poor representation within the CSD or PDB. Focusing on the CSD coverage for all cores found in five or more drugs results in a much improved average coverage of 91%. Unfortunately, the absolute number of drug molecular cores with low frequency is large. Similar to the observation by Bemis and Murcko,²⁸ this analysis found 73% of drug molecular cores to be in only one drug molecule. Still, 27% of these singleton drug molecular cores are found in the CSD.

Despite the small number of molecular cores present in PDB ligands, the drug molecular core coverage is relatively good. A total of 420 drug molecular cores are present in the PDB. As with the CSD, there is a strong correlation between frequency of occurrence in drugs and the probability of a structure being available in the PDB (Figure 2). As a

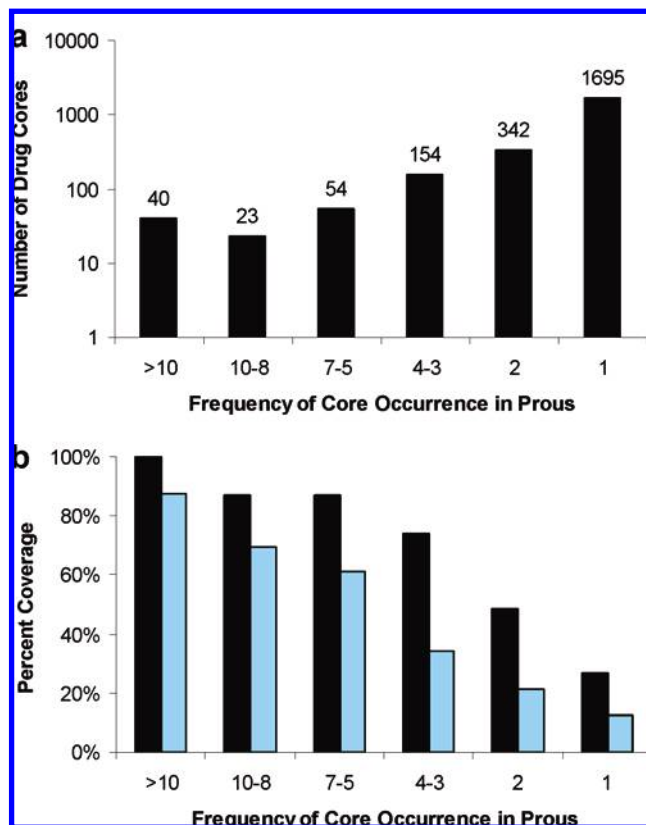


Figure 2. Frequency of occurrence of molecular cores found in compounds investigated in clinical Phase I trials and beyond. (a) The total number of unique drug cores for each frequency range. The largest category is for drug cores that occur in only one drug. (b) The fraction of drug cores observed in the CSD (black) and PDB (blue).

cautionary note, these estimated coverages must be considered upper limits because they do not consider the number of observations within the CSD or PDB for a given drug molecular core. For example, a drug molecular core observed only once in the CSD is still counted in this analysis. While such singleton coverage is of anecdotal value, it does not have statistical significance.

It is apparent that while the coverage of pharmaceutical small molecules in the CSD is not exhaustive, it is sufficient to be useful. Indeed, the chemical space visited with the highest frequency by medicinal chemists is well represented in the CSD. This corresponds to the chemical space we are most interested in studying from the perspective of conformation. As a consequence, the CSD has proven to be an excellent source of experimental data on the conformational preferences for common organic fragments used in medicinal chemistry.

The remainder of this article is organized as follows: We review preferred conformations of elementary acyclic systems (section 5) and sulfonamides (section 6). Due to their central role in medicinal chemistry, we discuss properties of aryl ring substituents in depth (sections 7 and 8). Here we also compare torsion angle histograms derived from CSD structures with those derived from ligands in the Protein Data Bank (PDB). This is followed by an analysis of biaryl systems and systems of two aryl rings connected by two acyclic bonds (section 9). We then discuss arylpiperidines and acylated piperidines as a special case of six-membered ring systems (section 10), give an overview of saturated and

unsaturated seven-membered ring systems (section 11), and close with a section on attractive sulfur–oxygen contacts (section 12). Since exhaustive coverage of structural elements relevant for drug design is beyond the scope of this article, we have focused on those systems whose properties go beyond what is discussed in introductory texts and which are highly relevant for drug design. In particular, the text covers more than half of the 20 most frequent substructures depicted in Figure 1.³¹

5. SIMPLE ACYCLIC SYSTEMS

The most fundamental organic fragments are saturated and acyclic. Saturated hydrocarbon chains such as butyl or longer, $\text{RCH}_2\text{CH}_2\text{—CH}_2\text{CH}_2\text{R}$, preferentially adopt a staggered conformation with the two terminal $\text{RCH}_2\text{—}$ groups *anti*. For butane, the most simple example of this class, steric repulsion between the C1 and C4 methyl groups (1,4-steric repulsion), increases the energy of the two *gauche* conformers by ~ 0.95 kcal/mol in the gas phase and ~ 0.55 kcal/mol in the liquid phase relative to the *anti* conformer.³² Within the CSD, this trend is extremely pronounced, and more than 90% of the C–C bond torsion angles of long alkyl chains are in an *anti* conformation ($170^\circ\text{—}180^\circ$). In addition to the intrinsic preference for an *anti* conformation, intermolecular crystal packing likely favors the extended geometry of this conformation which leads to a more compact crystalline structure. It would be expected that these packing effects are minimized for molecules with short alkyl chains. This is observed with a more restricted search of only terminal butyl chains, $\text{RCH}_2\text{—CH}_2\text{—CH}_2\text{CH}_3$, for which the prevalence of the *anti* conformation is reduced to 75% (Figure 3a shows data for this restricted search). As a molecular design element, unbranched alkyl chains are usually avoided due to their considerable flexibility. This property may be modulated and controlled with the introduction of branching substitutions, heteroatoms, or unsaturated linkages.

The dominant conformational feature in longer or branched hydrocarbons is the avoidance of *syn*-pentane interactions. For *n*-pentane, there are 9 possible staggered conformers that result from the combinations of the *gauche* and *anti* torsion angles for each of the two rotatable bonds. Symmetry operations reduce this ensemble to 4 energetically distinct conformers. Out of these, the conformer with a *gauche*+/*gauche*− sequence of central torsion angles (“*syn*-pentane”) is energetically highly unfavorable. A steric clash between the C1 and C5 methyl groups in *syn*-pentane gives rise to a calculated relative energy 3.3–3.6 kcal/mol above the lowest energy *anti*/*anti* conformation.³³ A CSD search of pentyl fragments, $\text{RCH}_2\text{—CH}_2\text{—CH}_2\text{R}$, confirms a complete absence of the *syn*-pentane conformation. *Anti*/*anti* and *anti*/*gauche* conformations are observed 91.6% and 7.5%, respectively, with the remaining conformations clustered near *gauche*+/*gauche*+ and *gauche*−/*gauche*−. The strong conformation preference away from *syn*-pentane has been used to design acyclic hydrocarbon backbones that are predicted to preferentially populate a single conformer, despite having up to six rotatable bonds.³⁴

The introduction of polar heteroatoms in unsaturated hydrocarbon chains can have a significant effect on conformation. For $\text{RCH}_2\text{X—CH}_2\text{CH}_2\text{R}$ butyl analogs where X = O or N, there is still a strong preference toward the *anti*

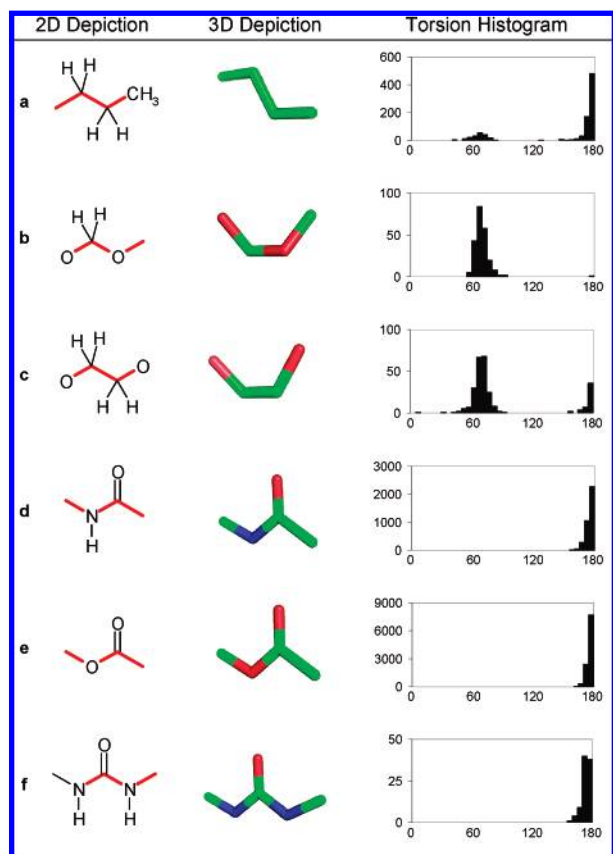


Figure 3. Torsion histograms for common acyclic organic fragments. The CSD searches included explicit hydrogen atoms where present in the 2D depiction.

conformation. The steric repulsion of the *gauche* conformers is expected to be slightly more pronounced than for the parent hydrocarbon due to the shorter $X-CH_2$ bond length.^{32,35} When searching the CSD, care was taken to ensure that the searches avoided any occurrences of an additional heteroatom within two bonds of the torsion of interest. The resulting data set is somewhat sparse, but 88% of the 120 observed fragments are in an *anti* conformation. The majority of these examples are short chains, so these results can be compared with the terminal butyl case which is 78% *anti*. The observed 88% *anti* population is consistent with a shorter C–O or C–N bond length, further disfavoring the *gauche* conformers.

The “anomeric effect” is an important stereoelectronic phenomenon that is present in $RY-CH_2Z$ systems where Y is an atom with one or more lone pair electrons and Z is electronegative. As a consequence of the anomeric effect, the common organic fragment, RCH_2O-CH_2OR , adopts a *gauche* conformation despite 1,4-steric repulsion. For a detailed discussion and review of the origins of this phenomenon, the reader is directed to ref 36. In our search of the CSD (Figure 3b), RCH_2O-CH_2OR is exclusively found as the *gauche* conformer, as previously observed.^{37,38} Substitution of the central methylene, as in $R^1O-CH(R^2)-OR$,³ introduces additional steric congestion, but the *gauche* conformer is still found in 71% of these cases, while the remaining 29% are evenly distributed between 100° and 170°. In these acyclic acetal systems, it has been observed that the second anomeric bond, $O-R^3$, often eclipses the anomeric hydrogen,³⁵ effectively leading to a distorted *syn*-pentane conformation. While less common as organic fragments and not well represented in the CSD except in

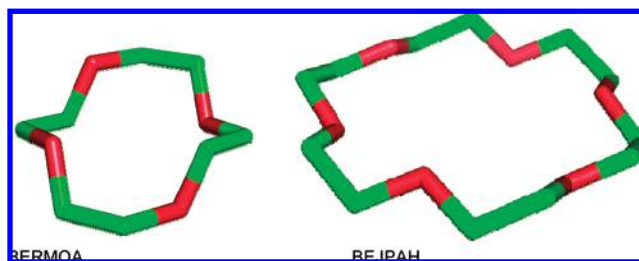


Figure 4. Two crown ether crystal structures with ring sizes of 12 (BERMOA) and 16 (BEJPAH) atoms. All the OC–CO bonds adopt a *gauche* conformation, while the remaining bonds are *trans*.

nucleosides, O–C–N and N–C–N systems also exhibit the anomeric effect^{39,40} as do thioethers.⁴¹ If such an N–C–N or O–C–N fragment occurs in a six-membered ring, alkyl substituents at N have a preference for the axial orientation caused by the anomeric effect. The anomeric effect is also observed with fluorine as a heteroatom (see section 8).

The propensity for ethylene glycol and 1,2-difluoroethane to prefer a sterically unfavorable *gauche* conformation was recognized by Wolfe and is commonly referred to as the “*gauche* effect”.⁴² The *gauche* effect has been widely studied and applies to all general systems XCH_2-CH_2Y , where X and Y are electronegative or have unshared lone pair electrons. A discussion of the theoretical explanations that have been put forth to account for this conformational preference is beyond the scope of this work but has been reviewed by Kirby.⁴³ The electronegativity of substituents X and Y is correlated with the degree of stabilization of the *gauche* conformer. Within the CSD, this trend is evident upon comparing the observed torsional distribution for cases where X and Y are either O or N. The *gauche* population for $ROCH_2-CH_2OR$ substructures is ~75% (Figure 3c). Substructures of type $ROCH_2-CH_2NR$ and $RNCH_2-CH_2NR$ show a strong preference to be *gauche* if one nitrogen atom is bound to an sp^2 carbon atom (anilines, amides), whereas alkyl bisamines of type $RNCH_2-CH_2NR$ are preferentially *trans* (no histograms shown). The strength of the *gauche* effect is readily apparent in crystal structures of crown ethers. Figure 4 shows the structures of two examples, BERMOA and BEJPAH, with ring sizes of 12 and 16 atoms, respectively. While these molecules have many rotatable bonds that adopt a *trans* geometry, all the $ROCH_2-CH_2OR$ torsion bonds are *gauche*. The *gauche* effect has also been investigated for systems involving fluorine.⁴⁴

The neurotransmitter acetylcholine (ACh, Figure 5) is a compelling example of the *gauche* effect in a bioactive molecule. On the basis of considerable steric crowding between the trimethylammonium terminus and the acetyl group, one may presume that a *gauche* conformation is disfavored. To the contrary, the small molecule structures in the CSD overwhelmingly adopt a *gauche* conformation, though the torsion distribution is broad and centered closer to 80°. The conformations of ACh in solution^{45,46} and bound to the catalytic site of acetylcholinesterase (AChE)⁴⁷ are also *gauche*, consistent with the crystallographic data. A *trans* conformation is reported in sterically hindered environments, such as in the peripheral anionic site of AChE⁴⁷ and the nicotinic ACh receptor.⁴⁶

Addition of a double bond to acyclic compounds provides considerable rigidity and a unique opportunity to control

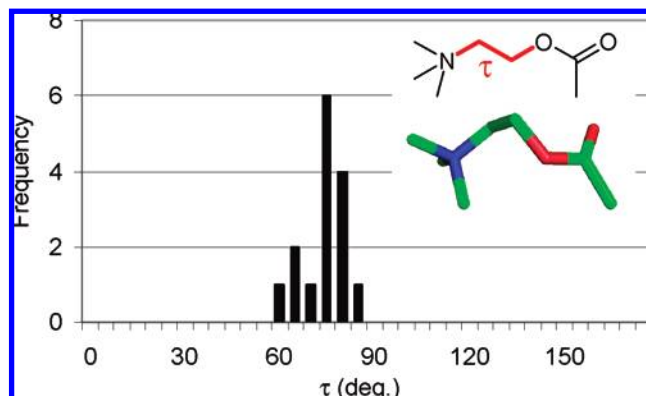


Figure 5. Torsion histogram for acetylcholine in the CSD. The *gauche* effect remains dominant despite significant steric crowding due to the trimethylamine terminus (in this case, the search constraint “no ions” was not used in ConQuest, as opposed to the other searches cited in the text).

conformation through the principle of allylic strain. For 1-butene, the central $\text{H}_2\text{C}=\text{CH}-\text{CH}_2\text{CH}_3$ bond prefers a conformation in which the $\text{C}=\text{CH}_2$ bond eclipses one of the $\text{C}-\text{H}$ bonds. Conformations in the CSD reflect this preference, and the torsion angle of the central sp^2-sp^3 bond has a skewed population centered at 120° with a minor *syn* population at 0° (Figure 6a). This minor *syn* population may be completely eliminated by branching at C3, as in 3-methyl-1-butene, without altering the skewed population centered at $120-130^\circ$ (Figure 6b).

Introduction of an additional substituent to an allylic center or Z to an allylic center leads to steric crowding and allylic strain. First proposed by Johnson and Malhotra,^{48,49} allylic 1,2- and 1,3-strain are widely applied in organic chemistry to exert stereoselective control in chemical transformations.⁵⁰ For a 1-butene fragment, substitution at C2 exerts allylic 1,2-strain on the adjacent sp^3 center. This is evident in the CSD torsion distribution which shows an increase in the *syn* population and a shift of the skewed population to $95-105^\circ$ instead of 120° observed for unsubstituted 1-butene (Figure 6c). An analogous steric condition exists in substituted aryl systems discussed in sections 7 and 8.

Allylic 1,3-strain occurs in the presence of a Z substituent on the double bond, such as a Z substituent at C1 of a 1-butene fragment. Any torsion angle of the central sp^2-sp^3 bond which brings a substituent on C3 in proximity to the C1 Z substituent will be sterically unfavorable. This 1,3-strain is not easily relieved given the rigid nature of the double bond. As a consequence, *syn* conformations are completely absent from the CSD for these molecules, and the torsion distribution is centered at $115-120^\circ$ (Figure 6d). In contrast to unsubstituted 1-butene, there is also a minor population of *trans* conformers. These are exclusively terminal $\text{CH}_3(\text{CH}_3)\text{C}=\text{CHCH}_2$ fragments, perhaps suggesting favorable packing for this group in a planar *trans* geometry. For highly substituted systems with both allylic 1,2- and 1,3-strain simultaneously, there are insufficient observations in the CSD to make any robust conclusions. However, the sparse distribution is consistent with an exclusion of *syn* and *anti* torsion angles, with most conformers lying between 90 and 120° .

A retrospective study of side-chain variants of mycophenolic acid provides a good example of the conformational

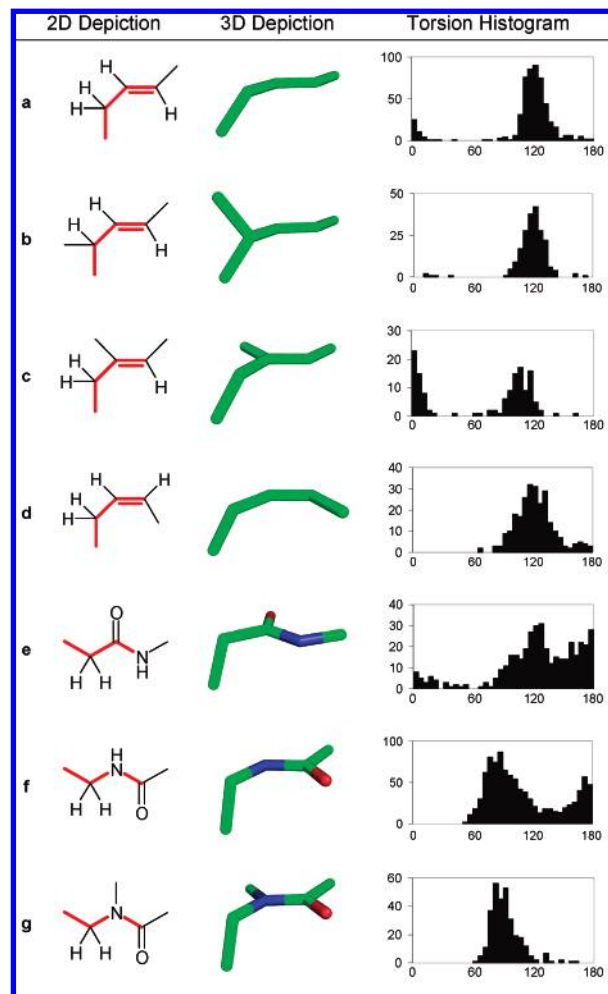


Figure 6. Torsion histograms for common allylic (or related) organic fragments. The CSD searches included explicit hydrogen atoms when present in the 2D depiction.

control that may be achieved in allylic compounds and its consequence on biological activity. Mycophenolic acid (**3**, Figure 7a) is an immunosuppressant that binds inosine monophosphate dehydrogenase (IMPDH) with an IC_{50} of $0.02 \mu\text{M}$. The X-ray structure of the complex has been solved (PDB 1jr1) and is shown in Figure 7b.⁵¹ The bioactive conformation of mycophenolic acid has an allylic torsion angle of $110-115^\circ$ which helps position the terminal carboxylic acid such that it can form two favorable hydrogen bonds in the binding site. The saturated hexanoic acid analog **1** is expected to be predominantly *anti* (180°) and has an IC_{50} of $1 \mu\text{M}$ against IMPDH. Improved potency of $0.14 \mu\text{M}$ is observed for the hex-2-enoic acid derivative **2** which is consistent with the predicted skewed torsion angle distribution centered near 120° , approaching the bioactive conformation. Finally, biasing of the conformation even further toward the bioactive bent structure is accomplished by introducing allylic 1,2-strain with a 3-methyl substitution, which will induce a torsion distribution centered around $95-105^\circ$. It is unlikely that this methyl group is contributing much hydrophobic binding energy as it is pointed toward solvent and makes little contact with the protein, yet it results in a 10-fold increase in potency. Conformational focusing is likely contributing significantly to the increased binding affinity. Figure 7a shows the distribution of torsion angles observed in CSD for these different organic fragments. While

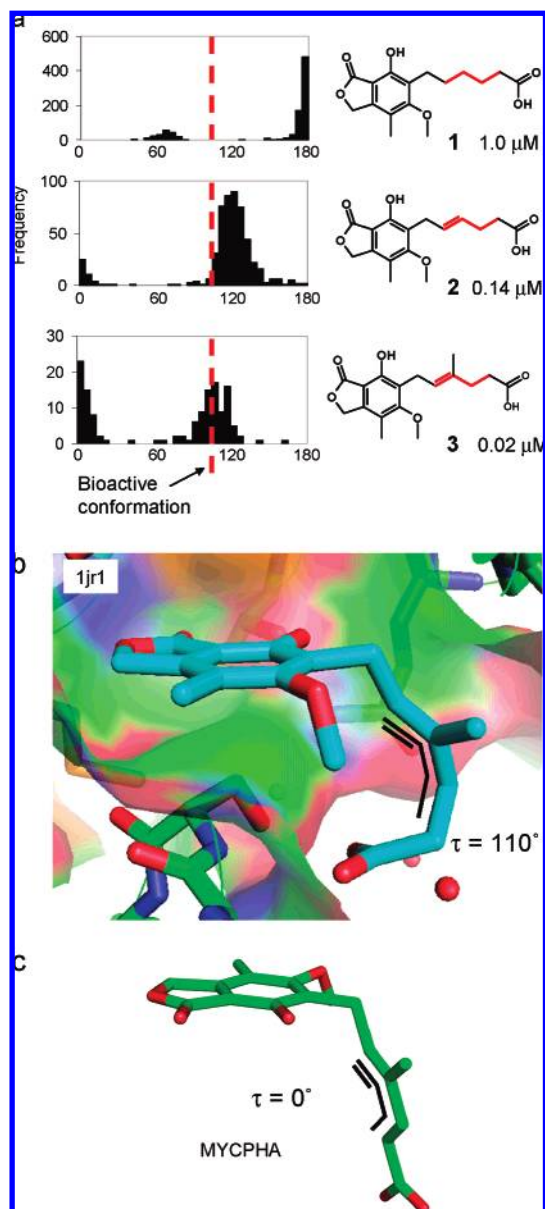


Figure 7. (a) Mycophenolic acid and two of its analogs. The torsion histograms for the allylic bonds highlighted in red are shown to the left. The red dashed line marks the observed torsion angle of 110° for the bioactive conformation. The conformation of mycophenolic acid (b) bound to IMPDH and (c) the small molecule structure in the CSD. The two structures adopt the two preferred geometries for systems under allylic 1,2-strain.

this is a retrospective study, it highlights the possibility of applying rational design to systems as simple as hydrocarbon chains. The mycophenolic acid case study also illustrates the benefits of using the complete CSD for conformational analysis. The small molecule crystal structure of mycophenolic acid deposited in the CSD (MYCPHA, Figure 7c) does not adopt the bioactive conformation. Instead, the bond under allylic 1,2-strain is *syn*, rather than skewed at 110 – 115° . If one were to rely exclusively on the single small molecule crystal structure of mycophenolic acid to build a model of the bioactive conformation, it would be incorrect. By searching the entire CSD for many diverse molecules that contain an allylic 1,2-strain functionality in different environments, it is evident that both the *syn* and skewed torsion angles are populated. Thus a model built from a complete

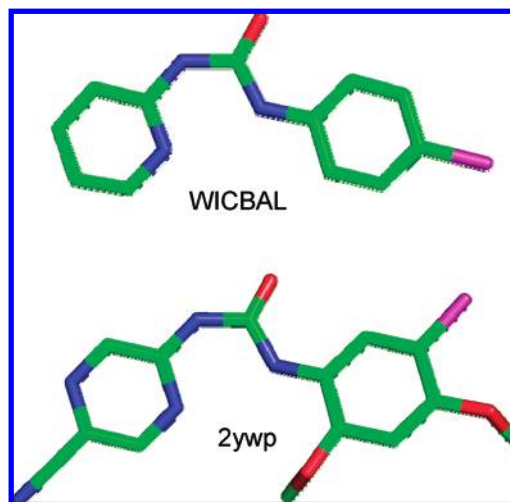


Figure 8. Two examples of *cis* amide conformations that occur as a consequence of intramolecular hydrogen bonding.

analysis of the CSD would correctly include both these possible side-chain conformations and potentially lead to a better target for molecular design.

Esters,⁵² amides⁵³ and their analogs, ureas, and carbamates are well-studied systems common in medicinal chemistry. They have a very strong preference to adopt planar conformations due to resonance stabilization. In fact, of almost 16 000 ester fragments in the CSD, only five anhydrides (and 4 structures that are most likely spurious entries) deviate more than 40° from planarity, and of more than 7500 amides, only 144 strongly sterically hindered tertiary amides have torsion angles between 40 and 140° . There are two possible conformations, *cis* and *trans*, for the amide or ester bonds. In acyclic secondary amides, esters, and ureas, the *trans* conformation is strongly favored, and, in the absence of special stabilizing interactions, the *cis* conformation is not observed (Figure 3d–f). For carbamates, 3.5% of the amide bonds have a *cis* conformation; however, the ester bonds always remain *trans* as has been previously observed.¹⁴ The hydrogen bond donor and acceptor pattern found in *cis* amides is a motif repeatedly used for binding to the “hinge” residues in the ATP site of protein kinases. By introducing an internal hydrogen bond, it is possible to stabilize the desired *cis* amide conformation in an acyclic linker. This strategy was applied successfully in the design of Chk-1 kinase inhibitors using 2-ureapyrazine analogs.^{54,55} The crystal structures of a 2-ureapyridine from the CSD (WICBAL) and a 2-ureapyrazine in complex with Chk-1 (PDB code 2ywp) confirm an intramolecular hydrogen bond between the pyrazine N1 and the urea NH atoms (Figure 8). The intramolecular hydrogen bond stabilizes one of the urea amide bonds in a *cis* conformation, allowing it to form the classic kinase-binding motif to the hinge region of the ATP site (see also section 8 for further details).

Varying the ring size of cyclic lactams and lactones provides an additional measure of the preference for *trans* versus *cis* amide and ester conformations. For cyclic lactams in the CSD, a transition is observed at a ring size of 9 atoms; lactams with a ring size smaller than 9 atoms contain *cis* amides, larger lactams have *trans* amides. The transition between *cis* and *trans* for esters in cyclic lactones occurs at

8 atoms. This suggests that esters have a stronger preference toward a *trans* conformation than amides and is consistent with the carbamate case previously discussed.

The strong resonance stabilization and planarity of amides and esters result in a phenomenon similar to allylic 1,2- or 1,3-strain. The carbonyl oxygen serves as the sterically offensive substituent. Thus functionalization adjacent to the carbonyl leads to pseudo 1,2-strain, and functionalization adjacent to the NH (or O of esters) leads to pseudo 1,3-strain. The carbonyl oxygen is not too large and does not extend out of the plane of the amide/ester bond; therefore, the steric effects are less pronounced than in a substituted allylic system. Nevertheless, the general trends are similar to allylic systems as can be seen by comparing the histograms in Figure 6c,d to those in Figure 6e (see also ref 56) and Figure 6f. For organic fragments with substituents adjacent to the NH of an amide, the *syn* conformation is completely absent, analogous to systems under the influence of allylic 1,3-strain. For tertiary amides, both 1,2- and 1,3-strain are present for substituents adjacent to the nitrogen. These systems show a strong avoidance of both *syn* and *trans* geometry with a focused population between 80 and 110° (Figure 6g). Note that depending on the substitution pattern, electronic effects may play a role as well. For example, it has been noted that α -fluoroamides show a strong preference for a *trans* arrangement of the O=C–C–F unit.⁵⁷

A parallel study was completed for all the acyclic torsions discussed above using protein-bound ligand conformations deposited in the PDB. In general, the PDB histograms show the same trends as observed in the CSD, but with considerably broader distributions. These histograms are included in the Supporting Information and do not change any of the conclusions drawn from the CSD analysis. A broader conformational distribution for PDB ligands is consistent with the recent study by Hao et al.⁵⁸ in which *ab initio* DFT torsion potentials were compared to torsion profiles of PDB ligands for 21 common torsion motifs. A Boltzmann distribution derived from these calculated torsion potentials required a temperature factor much higher than room temperature to reproduce the observed PDB observations. This was rationalized as protein induced structural perturbations resulting in an average strain energy of ~ 0.6 kcal/mol/torsion. We observe the greatest discrepancy between the CSD and PDB derived torsion histograms to occur for the torsion motifs with the softest energy potentials such as the OCCO gauche effect and allylic systems. Model bias and experimental noise also may contribute to the differences between CSD and PDB distributions. These phenomena are discussed in section 7.

6. SULFONAMIDES AND SULFONES

The first CSD study of arylsulfonamides was published in 1986 by Beddoes et al.⁵⁹ At the time, X-ray structures of 49 acyclic sulfonamides were available. The conclusions drawn on the lowest energy conformation have not changed (Figure 9). A simple mnemonic rule can be derived from the Newman projections along the N–S and the C–S bonds: Both the nitrogen lone pair and the carbon aromatic p orbital bisect the O=S=O angle. It has been suggested that these rotamers are stabilized by lone pair interactions with sulfur d orbitals.⁶⁰ A state-of-the-art investigation which

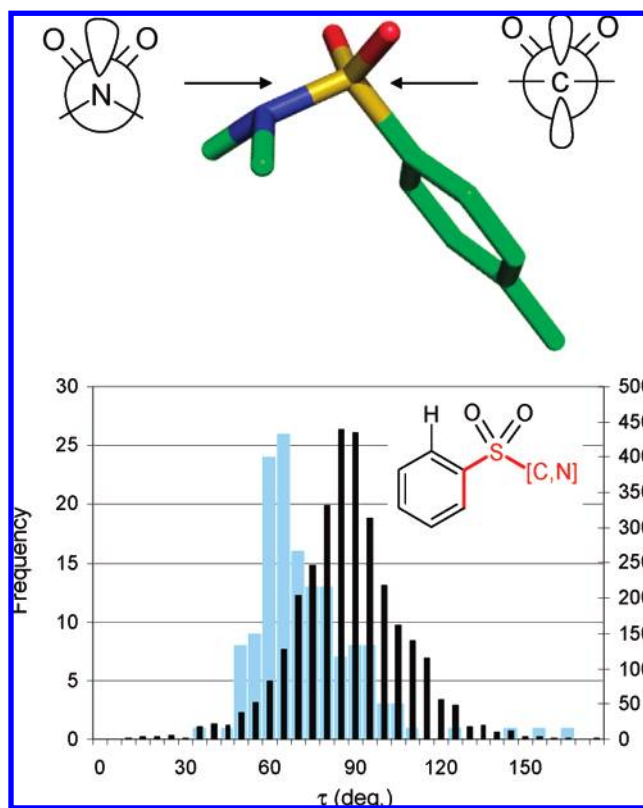


Figure 9. Top: An exemplary aryl sulfonamide structure from the CSD (GESSUS) and Newman projections for the two rotatable bonds. As a rule, sulfonyl compounds adopt conformations in which the substituent lone pair (here the N lone pair or the p orbital at the *ipso* carbon) are bisecting the O=S=O angle in the Newman projection. Bottom: Overlays of two CSD torsion histograms for aryl sulfonyl motifs without *ortho* substituent (black, X=H) and with one *ortho* substituent (gray, X≠H).

could deliver a more detailed explanation is still missing. Whatever their origin, the conformational preferences of sulfonamides are actively exploited in drug design. For example, a group at GSK recently described conformational preferences in sulfonamide Factor Xa inhibitors including *ab initio* calculations.^{61,62}

The number of aryl-sulfonyl structures in the CSD is now 2 orders of magnitude larger than 20 years ago, allowing a more detailed analysis. Figure 9 shows torsion histograms of aryl-sulfonyl single bonds. In the absence of *ortho* substituents, the expected peak at 90° is observed. The torsion distribution is rather broad and approximately of the same shape as the torsion histogram of primary alkyl substituents attached to aryl rings (see also Figure 12), indicating a rather soft local potential. As a general rule, the two sulfonyl oxygen atoms are situated on one side of the aryl ring (this is the case for torsion angles between 60° and 120°). In the presence of one non-hydrogen *ortho* substituent, the peak of the distribution shifts to about 70°, i.e. the sulfonyl group rotates away from the substituent until one S=O bond is almost coplanar with the aryl ring. Figure 11 contains one typical example (QIBRIC) with a methyl group in the *ortho* position. Note that the second phenyl group is also in plane with one S=O bond to avoid 1,6-repulsion with the methyl substituent on the other ring.

An overview of sulfonamide conformations requires the analysis of the corresponding S–N torsion angle as well as the hybridization of the nitrogen atom. We use the distance

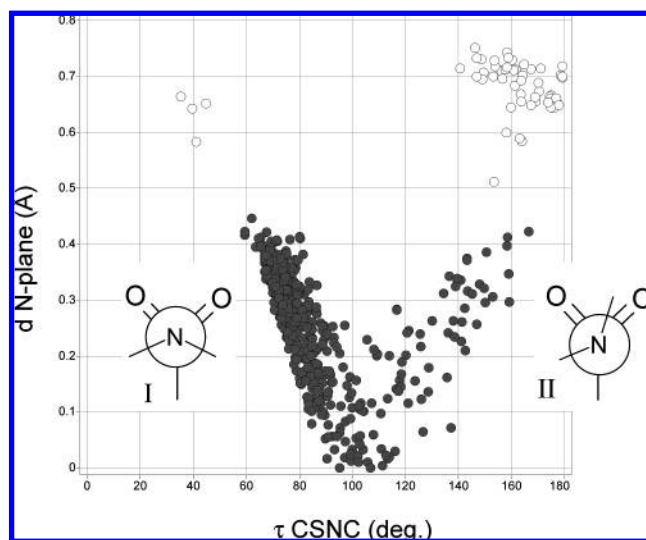


Figure 10. Scatterplot for *N,N*-disubstituted sulfonamides in the CSD. The distance between N and the plane of its three substituents is plotted against the C–S–N–C torsion angle. The absolute value of the larger of the two alternative C–S–N–C torsion angles was chosen to distinguish between the two basic types of conformations. The left-hand branch in the plot corresponds to the strongly preferred standard conformation depicted in Figure 9.

between the nitrogen atom and the plane spanned by its three substituents as a measure of pyramidal-ity. For acyclic tertiary amines, this distance is typically around 0.45 Å. Sulfonamide nitrogen atoms are generally slightly less pyramidal with an N-plane distance between 0.25 and 0.4 Å but often also significantly lower. In Figure 10, the N-plane distance is plotted against the C–S–N–C torsion angle in *N,N*-disubstituted sulfonamides. For each CSD structure, two such torsion angles can be defined, since there are two carbon substituents on the sulfonamide N. For the plot, the absolute value of the larger of the two torsion angles was chosen to distinguish between the two basic types of conformations found. The vast majority of the structures adopt the “lone pair bisects O=S=O angle” conformation described above (Figure 9). In this case, the C–S–N–C angle is generally less than 90° and decreases with increasing N pyramidal-ization. A second, far less populated cluster of structures is formally related to the first by a nitrogen inversion, whereby

one N substituent and the N lone pair change positions (Newman plot II in Figure 10). In this case, there is a positive correlation between increasing pyramidal-ity and the torsion angle. Closer examination reveals that this second conformation is only adopted by structures with branched N substituents, which would suffer significant steric strain in the first conformation. Typically, these are derivatives of α-substituted cyclic amines such as ADISAH (Figure 11). Conformation I should thus be significantly lower in energy than II. Interestingly, very large basis sets are required to reproduce any significant energy difference between the two conformations in model systems. Model calculations on geometry-optimized *N,N*-dimethylmethanesulfonamide at the MP2/cc-pVDZ or the B3LYP/cc-pVDZ level give no energy difference between the two local minima I and II. Using a larger basis set of cc-pVTZ quality energy differences of 1.1 and 0.9 kcal/mol (II–I), respectively, are calculated, which are likely to be a lower bound for the experimentally observed preferences.

The scatter plot in Figure 10 reveals one more interesting fact. The open circles with small absolute values of $\tau(\text{SCNC})$ (i.e., with the most acute bond angles at N) are sulfonyl derivatives of aziridines. In a striking reversal of the conformational preference observed above, these structures preferably adopt conformations of type II. An example is given in Figure 11 (KATLEW). Only four exceptions exist in the CSD (AVAKOX, BONMAS). It seems as if in these cases the aziridine Walsh orbital⁶³ plays the role of the N lone pair; it is situated such that it bisects the O=S=O angle in a Newman projection.

The general conformational preference of Figure 9 is equally valid for bis-sulfonamides,⁶⁴ aliphatic sulfonamides, and sulfamides, and it has two important practical consequences: First, cyclic sulfonamides preferentially carry axial substituents, since only this arrangement corresponds to the preferred rotamer. This is generally the case in the CSD for cyclic sulfonamides in six-membered and larger rings (e.g., HITQOQ, Figure 11), except for those where other steric effects override this preference (YAKBOC, XOHBOL, not shown). Smaller rings are too strained to adopt the optimal S–N rotamer and tend to have equatorial substituents (few examples in the CSD). A second feature of sulfonamides is their ability to form bent structures in which two substituents,

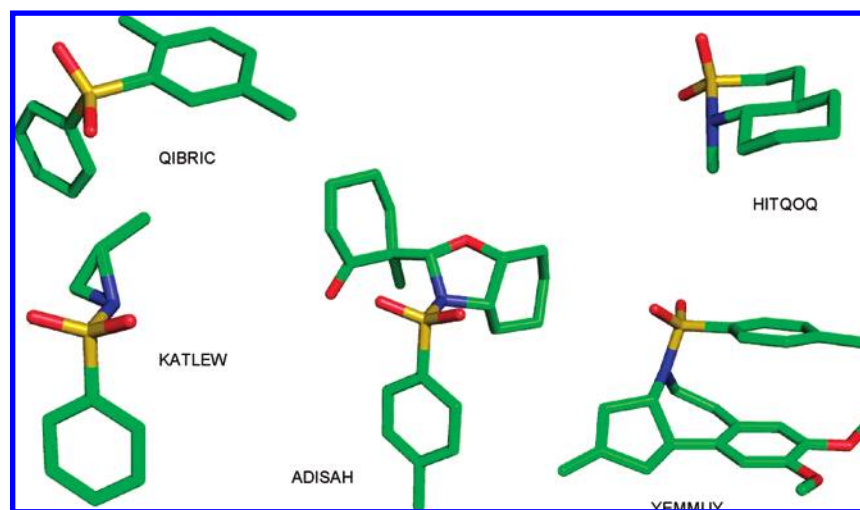


Figure 11. Various CSD structures illustrating structural properties of sulfonyl groups (details see text).

typically ring systems, fold back onto each other. This is also favored by the larger size of sulfur, forming longer bonds to second-row elements and making 1,5-interactions, strongly avoided by carbon skeletons, much less repulsive or even attractive. The structure YEMMUY in Figure 11 is an example. The PDB contains many further examples; in particular there is a large class of matrix metalloproteinase inhibitors exploiting this principle.

7. SIMPLE SUBSTITUTED ARYL SYSTEMS

This section serves a 2-fold purpose. We review the rules governing conformational properties of some frequently occurring aryl substituents and at the same time compare torsion histograms generated from CSD structures and small molecules contained in the PDB.

While detailed comparisons of molecular *interaction* patterns found in the PDB and the CSD have been published,⁶⁵ surprisingly little work has been done in comparing *conformational* preferences of typical substructures found in the two databases. A study comparing conformations of individual molecules occurring in both databases led to the conclusion that these structures do differ significantly.⁶⁶ Other studies focused on comparisons between PDB ligand structures and calculated (vacuum or solution) global minima conformers,^{67,68} and a recent study came to the conclusion that ligand strain energy is not correlated with binding affinity.⁶⁹ Here we are not interested in explicit conformational energies but in differences in observed torsion angle distributions derived from the two databases and the potential causes of these differences.

Figure 12 shows torsion histograms of six frequent aryl substituents with hydrogen atoms located at the *ortho* positions. In each case, the CSD distribution (black) is overlaid with the one found in the PDB (light blue). Overall, CSD and PDB preferences match well, with the PDB profiles showing less clear-cut preferences. There are two potential causes for these broader distributions: the lower coordinate precision of protein X-ray structures or a different *a priori* behavior of small molecule ligands in protein environments, i.e., a higher degree of adaptation of ligand conformation in the protein than in small molecule structures. Also, protein complex structures usually contain a large number of water molecules, while small molecule structures are typically crystallized from apolar solvents. The presence of a high dielectric solvent may alter the statistics of protein-bound ligand conformations compared to CSD data, in particular when intramolecular hydrogen bonds are stabilizing a particular conformation.

Although the relative importance of these two factors cannot be assessed here, there are strong indications that the higher degree in uncertainty in assigning exact coordinates to PDB structures plays a major role.⁷⁰ Coordinates of ligands in protein X-ray structures are generally refined by weighting the density map against a model structure of the ligand. These models are typically created with a rule-based 3D structure generator or minimized with standard molecular mechanics force fields. As a result, model artifacts are carried over into the refined coordinates.

A particularly striking case in point is the profile of benzamide structures (Figure 12a). The CSD torsion profile peaks at angles of around 30° and 150°. Ab initio

calculations confirm that the fully planar arrangement is a local maximum, not a minimum of the torsion profile.^{71,72} In contrast, the majority of the PDB structures are fully planar. Standard force fields typically generate fully planar benzamide conformations, and in addition thermal averaging and the low resolution of protein structures will often prevent the determination of a preference for one of the out-of-plane positions. For benzanilides (Figure 12b), the situation is subtly different. These systems do adopt a planar lowest energy conformation,⁷³ and this structure is also found most frequently in the PDB, but deviations from it are much less frequent than in the CSD—again, most likely due to a model bias (note that acetoxyphenyl systems, the oxygen analogs of *N*-phenylacetamides, adopt almost orthogonal conformations with torsion angles between 60° and 120°⁷⁴).

The torsion distribution of all aryl-sulfonyl systems in the PDB is in perfect agreement with the one derived from the CSD (Figure 12c). A large majority (85%) of the structures represented by this histogram are *N*-substituted sulfonamides. The remaining structures are primary sulfonamides or methyl sulfones. Figure 12d shows the torsion angle distribution for this subgroup of sulfonyl substituents: no conformational preference is observed, while the CSD distribution in Figure 12d remains the same as that for the full set of aryl sulfonyl structures, albeit at lower statistical significance. This striking difference is most certainly caused by the fact that in electron densities of protein–ligand complexes, C, N, and O positions are typically not distinguishable.⁷⁵ For primary amides and sulfonamides, an analysis of intermolecular contacts is required to determine the most probable atom assignment. A case in point are the two CDK2 ligand structures depicted in Figure 14, where both sulfonamides have approximately the same protein environment but different rotamers have been deposited.

The distribution of alkoxy group torsions in the PDB differs most strongly from that found in the CSD (Figure 12e). Here, the CSD distribution has a sharp maximum at 0°, indicating a strong preference for a planar orientation as the only minimum in agreement with high level *ab initio* calculations.^{76–78} An earlier CSD study by Bürgi *et al.*⁷⁹ lists a number of structures with angles around 90°, which are most likely cyclic derivatives. Here we included structures with acyclic aryl-O bonds only and do not observe perpendicular orientations. In the PDB, the preference for a planar orientation is also observed but it is much weaker and there are a significant number of structures with angles around 90°. The occurrence of nonplanar structures in the PDB is very high even for methoxy groups, for which strain caused by packing effects should be minimal (Figure 12f). As the *ab initio* calculated torsional potential energy of anisole^{76,78} continuously increases from its planar minimum to a barrier height of approximately 3 kcal/mol in the orthogonal conformation, we suggest that the majority of the nonplanar structures are fitting artifacts. Also, spectroscopic studies on anisole–water complexes indicate that anisole stays planar when interacting with water monomers, dimers, or trimers.⁸⁰ A detailed analysis of many individual complexes will be required to answer the general question if protein environments do or do not induce deviations from planarity more often than small molecule crystal environments.

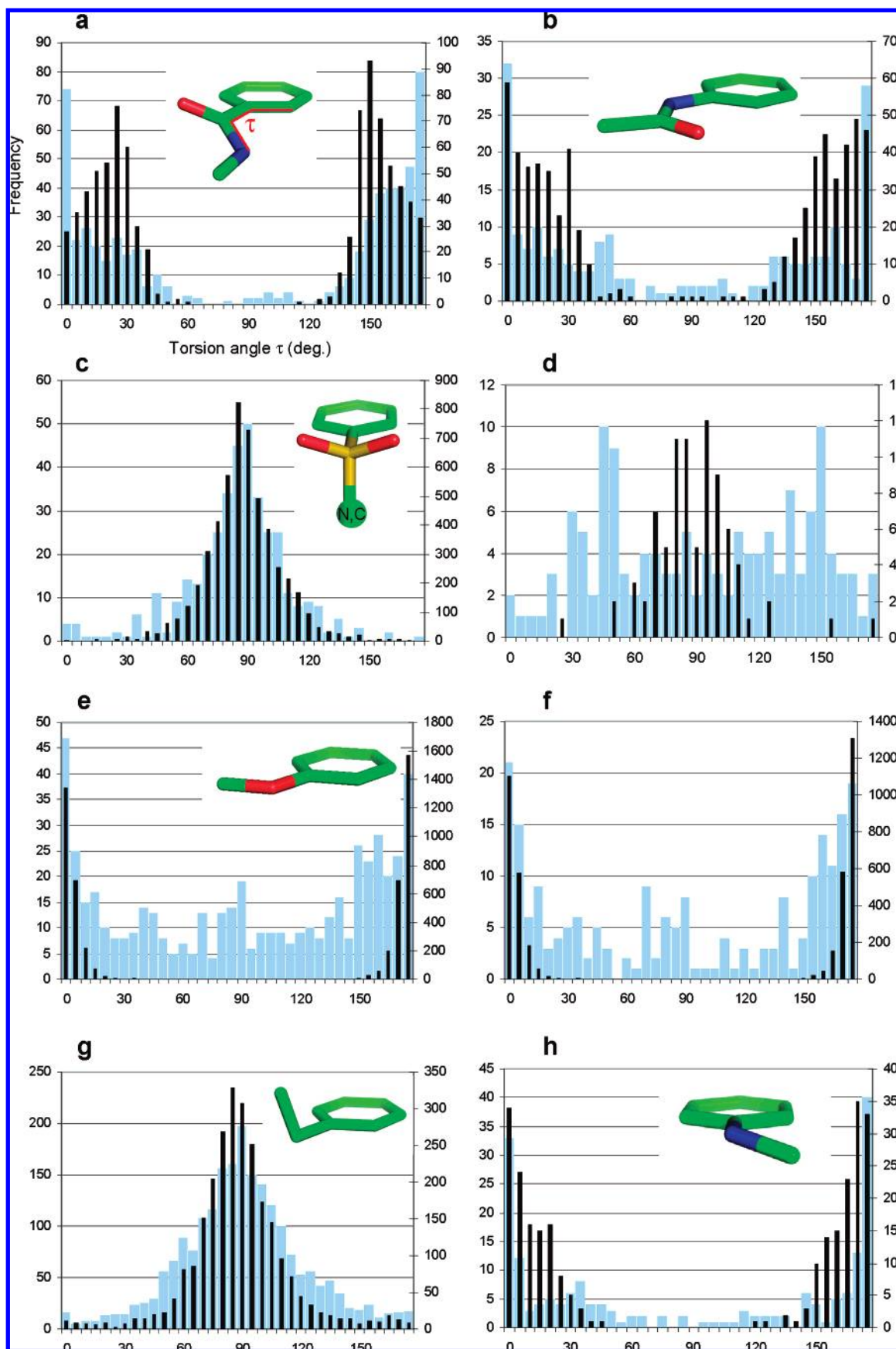


Figure 12. Torsion histograms (occurrence vs torsion angle bins) for the most frequent aryl substituents in the absence of ortho substituents. In each case, the PDB histogram (light blue, y-axis to the left) is overlaid with the corresponding histogram derived from CSD ligands (black, y-axis to the right): (a) primary benzamides, (b) acylated anilines, (c) aryl sulfonyl fragments (sulfonamides or sulfones), (d) a subset of (c) which are either methyl sulfones or sulfonamides carrying no N substituents, (e) primary alkoxy substituents, (f) a subset of (e) where methoxy substituents are omitted, (g) primary alkyl substituents, and (h) N,N-disubstituted anilines.

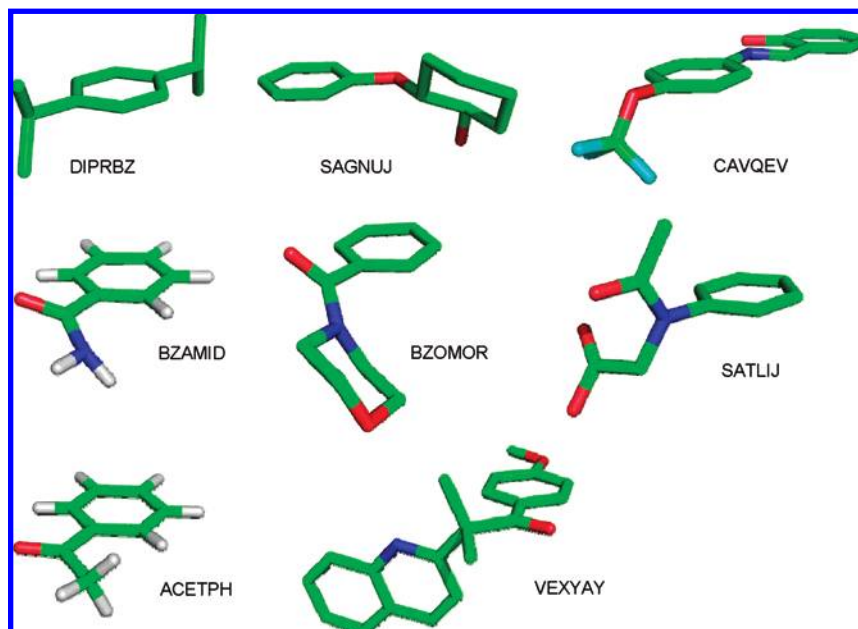


Figure 13. Various CSD examples of sterically more demanding aryl substituents (details see text).

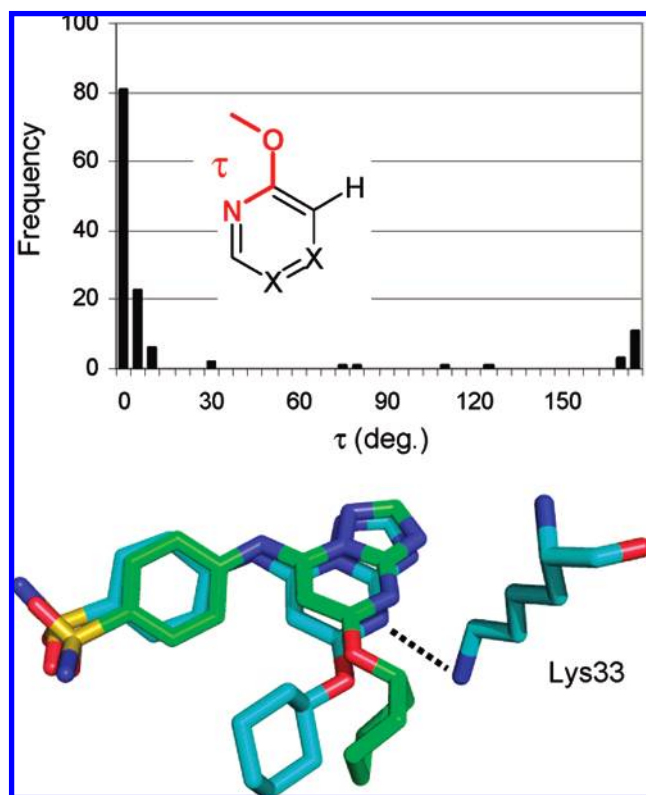


Figure 14. Top: Torsion histogram for *ortho*-alkoxy pyridines illustrating the strong avoidance of the *anti* orientation. Bottom: Overlay of two protein–ligand complexes containing the same inhibitor NU6102 with monomeric CDK2 (2c6t, green) and with the CDK2/cyclin A complex (2c6m, blue). In the CDK2/cyclin A complex, the nitrogen lone pair forms a hydrogen bond to Lys 33 and the neighboring substituent can adopt an *anti* orientation.

Figure 12g shows the torsional preference for primary alkyl substituents. In general, these tend to assume a perpendicular orientation to the aryl ring. This conformation may be seen as a “compromise” between avoiding steric repulsion (akin to allylic 1,3-strain) with the aryl CH groups in either *ortho* position. The broadness of the distribution is indicative of a

rather soft torsion potential, which will become apparent again in the section on aryl-X-aryl systems below.

To complete the picture of typical aryl substituents, the torsion histogram for N-monosubstituted anilines is depicted in Figure 12h. The general tendency for almost planar structures is clearly apparent, and there is good agreement between the PDB and CSD torsion distributions. It has to be noted, however, that the simple torsion profile cannot do justice to the actual conformational properties of aniline systems. Depending on the environment and the electronic nature of additional ring substituents, nitrogen hybridization can vary considerably. Anilines are often slightly pyramidal, with the larger N substituent (in this case the alkyl substituent) pointing out of plane. This is reflected in the CSD histogram but not in the PDB-derived one. Dimethylaniline has a nonsymmetrical minimum structure (REDDAF, not shown) with one NC bond in plane with the ring and the other one deviating 14° from the ring. This asymmetry is relevant for very common motifs in medicinal chemistry such as arylpiperidines discussed in section 9.

8. SUBSTITUTED ARYL SYSTEMS – STEREOELECTRONIC AND STERIC EFFECTS

Thus far we have discussed conformation preferences of aryl ring substituents in an “unperturbed” state, that is, in the absence of *ortho* substituents, heteroatom effects, or special steric constraints. Here we will not exhaustively cover combinations of substituents and their effects. In particular, we will not discuss the steric influence of *ortho* substituents. These typically shift the equilibrium toward one side, as illustrated above for aryl sulfonyl groups (Figure 9), or block one rotamer in the case of two alternative in-plane rotamers (e.g., for methoxy groups). Here we focus on selected properties of sterically demanding aryl substituents and subsequently demonstrate some stereoelectronic effects.

Any of the aryl substituents discussed in the previous section can occur in a more highly substituted and thus sterically more demanding form. Under some circumstances a higher degree of substitution leads to altered conformational

preferences. This is particularly true if the preferred conformation of the unperturbed system is planar and additional substituents would lead to steric clashes within the plane of the aryl ring. All CSD structures referred to in the following paragraph are depicted in Figure 13. No histograms are shown.

While primary alkyl substituents prefer a 90° orientation, secondary alkyl substituents are typically positioned at 60° relative to the plane of the ring, above and below the ring plane (e.g., as in DIPRBZ), in a way analogous to the preferred conformation of 3-methyl-1-butene. Secondary alkoxy groups and even tertiary ones have almost the same strong tendency to be planar as primary alkoxy groups. Note that secondary systems such as SAGNUJ have two locked rotatable bonds, because rotamers pointing one of the two alkyl branches toward the ring are avoided. An unusual change in the planar preference of alkoxy groups occurs for the trifluoro-alkoxy group in which the aryl-O bond prefers to be perpendicular to the ring plane. With only 16 examples in the CSD and PDB combined, the prevalence of this group is rather small to draw robust conclusions. However, these few observations strongly favor a perpendicular geometry as exemplified in structure CAVQEV. This perpendicular preference is in good agreement with *ab initio* calculations that indicate a global minimum geometry at an aryl-O torsion angle of 90°. ⁷⁸ The preference for an orthogonal arrangement for R = CF₃ vs planar for R = CH₃ can be rationalized by (i) the larger volume of the trifluoromethyl group leading to higher 1,4 strain in the planar conformation and (ii) electronic effects of the highly electronegative CF₃ group reducing the amount of π -conjugation of the oxygen ether atom with the aryl system (anomeric effect between oxygen lone pairs and C–F bonds).

We have seen above that primary and secondary benzamides are typically bent out of plane by up to 30° (BZA-MID). Tertiary benzamides, in contrast, have wider torsion angles between 30° and 60° to avoid steric clashes between the aryl ring and the amide *cis* substituent (BZOMOR). In a similar fashion, acetylated aniline derivatives carrying a second substituent at the aniline N are generally strongly twisted structures (SATLIJ).

As opposed to benzamide, acetophenone and other primary aryl ketones have fully planar global minimum structures. The fully planar acetophenone structure becomes possible most likely because the hydrogen atoms of the methyl group are positioned above and below the plane of the aryl ring (ACETPH) instead of lying in the plane of the ring as would the amide NH in the case of a planar benzamide structure. Interestingly, the keto group is typically not strongly twisted out of plane even in quaternary aryl ketones. Torsion angles between 0° and 30° are the norm for these systems, e.g., as in VEXYAY.

Lone pair repulsion can have strong effects on conformation. 2-Alkoxy pyridines are excellent examples to study this phenomenon. The histogram for the torsion angle between the ring and the alkoxy group shows a clear preference for the *syn* orientation. In the *anti* orientation, the N and O lone pairs would repel each other. Those CSD structures that do show a deviation from the *syn* rotamer invariably display an intermolecular hydrogen bond to the pyridine N in the crystal environment. In these cases the N lone pair is “occupied” and no longer enforces the conformation of the

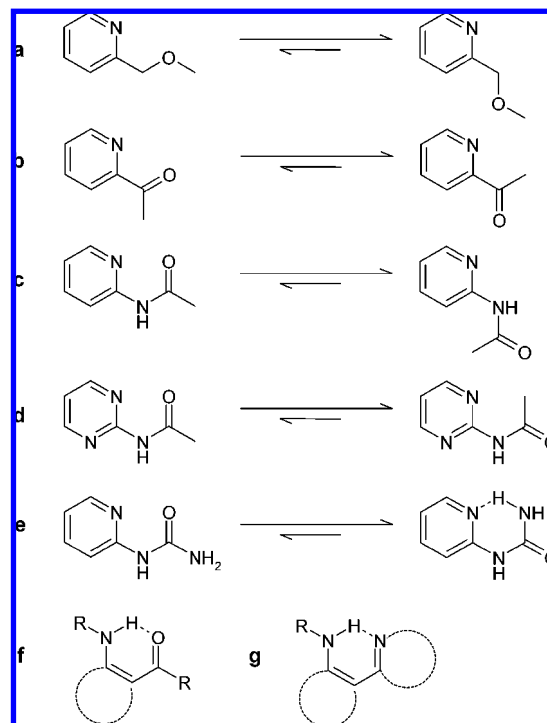


Figure 15. Structural elements adopting preferred conformations due to lone pair repulsion (a–d), intermolecular hydrogen bonds (f, g), or a combination thereof (e).

neighboring substituent. In this sense one can think of the hydrogen-bonding partner as an additional substituent. The PDB contains one particularly striking example that illustrates this effect. A group at Vernalis studied triazolopyrimidine derivatives as inhibitors of CDK2.⁸¹ They report the complex structures of the same inhibitor NU6102 with monomeric CDK2 and with the CDK2/cyclin A complex. Both structures are deposited in the PDB and depicted in Figure 14. Only in the CDK2/cyclin A complex (2c6m), the nitrogen lone pair forms a hydrogen bond to Lys 33 and the neighboring substituent can adopt an *anti* orientation. In the other structure (2c6t), such a hydrogen bond is not formed, and the cyclohexyl substituent switches to the *syn* rotamer, inverting to the axial form to adapt to the binding pocket. Interestingly, the authors did not comment on this remarkable finding.

There are a variety of analogous cases to *ortho*-alkoxy pyridines: Pyridines substituted with alkoxyethyl groups in *ortho* position avoid the *syn* rotamer (Figure 15a). Here the *syn* rotamer would lead to strong electrostatic repulsion, and the *anti* rotamer is in addition slightly stabilized by a CH...O interaction. Thus this system has an intrinsic conformational preference opposite to that of 2-alkoxy pyridines. The same holds for pyridines carrying an acyl function in the *ortho* position (Figure 15b) as well as for acylated aminopyridines (Figure 15c). In particular for this latter case, CSD statistics strongly support the indicated preference. The situation is slightly different for acylated 2-aminopyrimidines (Figure 15d). Both alternative in-plane rotamers of the acylamino group would suffer from the same repulsive interaction between C=O and N. Thus, these systems typically adopt a *cis* amide conformation, which is normally a rare event as discussed in section 5. The preference for a *cis* amide conformation is even stronger if it is stabilized by an intramolecular hydrogen bond (Figure 15e). However, when

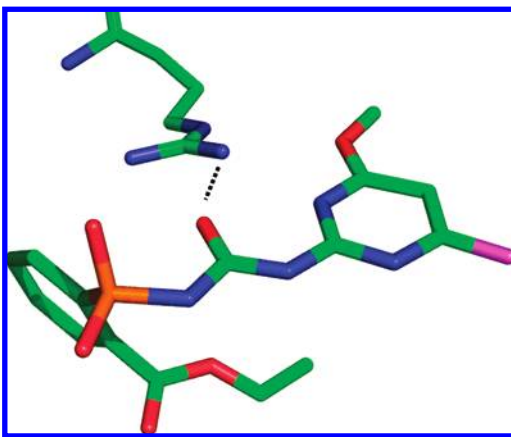


Figure 16. Example of a PDB structure (1ybh, acetolactate synthase complex) where the structural element depicted in Figure 15e adopts a *trans* amide conformation due to the formation of an intermolecular hydrogen bond (with the side chain of Arg 377).

analyzing alternative conformations of this kind, the environment of the system must always be considered. For structures of the type depicted in Figure 15e, the presence of a strong hydrogen bond donor in the protein can revert the conformational preference to the *trans* amide. The PDB contains several examples where a lysine or arginine side chain fulfills this role, as exemplified by 1ybh (Figure 16).

To complete the picture, Figure 15f,g shows two generic substructures in which the intramolecular hydrogen bonds formally form a six-membered ring. Due to their geometry, these are the most stable hydrogen-bonded systems. These and related structural elements have been used to design a wide variety of rigidified templates, e.g., scaffolds of inhibitors for protein–protein interactions.⁸² The analogous five-membered ring systems are also found but are less relevant in forming stable structural elements. Finally, it should be mentioned that these classical hydrogen bond systems have weak counterparts in which a CH unit, typically in an electron poor heterocycle, takes the role of the hydrogen bond donor.⁸³ Such bonds play a role in intermolecular interactions, as has become particularly apparent in the binding modes of kinase ATP site inhibitors.⁸⁴ If they occur intramolecularly, they make planar conformations possible where they would not be tolerated otherwise due to steric constraints. As an example, diarylamines are normally twisted in a fashion similar to the benzophenones discussed in section 8. N-Aryl substituted 2-aminopyrimidines and related structures can adopt fully planar structures (e.g., PDB entries 2np8, 1rww, not depicted).

9. BIARYL AND ARYL-X-ARYL SYSTEMS

Biaryl fragments are common in medicinal chemistry and can adopt a wide range of conformations depending upon substitution patterns and ring sizes. Biphenyl is the most common drug molecule core (see section 3) and well represented in the CSD. In the absence of *ortho* substituents, the central C–C bond is twisted to give an angle of 30–40° between the two ring planes. A second peak in the plane angle histogram at 0° accounts for ~20% of the CSD structures and has been shown to be the result of crystal packing.¹² Comparison of the PDB and CSD histograms show a tight correlation for the major peak at 30–40°;

however, the crystal packing artifact at 0° is clearly absent from the PDB histogram (Figure 17a). Introduction of an *ortho* N, as in 2-phenylpyridine, reduces the prevalence of the anomalous planar conformations to only 7% while globally shifting the average plane angle to ~15°. A second N, as in 2-phenylpyrimidine, allows the aromatic rings to be nearly coplanar with an average plane angle of ~5°.

Five-membered N and O heterocycles have less steric crowding and, when substituted with a phenyl ring, prefer to place both rings coplanar. While crystal packing effects cannot be ruled out, the observed histograms for the CSD and PDB show a reasonable correlation with no anomalous peaks (Figure 17c). One may hypothesize that the larger ring size of thiophene would adopt a plane angle in the range between the smaller coplanar five-membered heterocycles and biphenyl. This is indeed observed, and 2-phenylthiophene populates a broad distribution of plane angles <30° with two minor peaks at 5° and 25°. In this case, crystal packing forces likely account for the coplanar observations, and gas-phase calculations predict a preferred ring plane angle of 25°.⁸⁵

Any substitution *ortho* to the phenyl ring in biphenyl eliminates all conformations with coplanar rings and increases the average ring plane angle to 55° (Figure 17b). A second *ortho* substitution on the same ring introduces even more steric crowding, and the average ring plane angle is 72°. A similar trend is observed for 5,6 biaryl systems when the phenyl ring bears an *ortho* substituent. If the five-membered ring is *ortho* substituted, then the angle between the ring planes increases from <30° to a broad range between 15° and 50°, with an average at 35° (Figure 17d). This is still considerably less than the ring plane angle for *ortho* substituted biphenyls. These examples show that if a biaryl torsion angle is a critical design component, it may be carefully controlled by modifying the nature of the rings and their substituents.

A sizable number of drug molecules contain two six-membered aromatic rings linked by two acyclic bonds via a simple functional group such as methylene, carbonyl, NH, or a sulfur or oxygen atom. In general medicinal chemistry practice, these linkers are regarded as isosteres and are often interchanged, but they have quite distinct preferred conformations and degrees of flexibility. Partly, these differences can be deduced by considering these linkers to be the simple aryl substituents as outlined above. However, the additional steric interplay of two closely neighboring ring systems warrants a more detailed analysis of the two consecutive dihedral angles determining the conformations of these systems.

The plots in Figure 18 are composite views of torsion angle distributions derived from the CSD (scatterplot) and energy hypersurfaces derived from AM1 semiempirical calculations (contour lines in distances of 0.1 kcal/mol). The semiempirical hypersurfaces are in excellent agreement with the CSD statistics, and the contour lines were included because they give a more vivid impression of the low-energy regions of the torsion profiles. For diphenylmethane, the global minimum is a structure where the two phenyl rings are positioned orthogonal to each other, as one would expect from the univariate torsion profile of an aryl ring with primary alkyl substituents. However, most structures deviate strongly from this minimum energy structure. Many structures are located

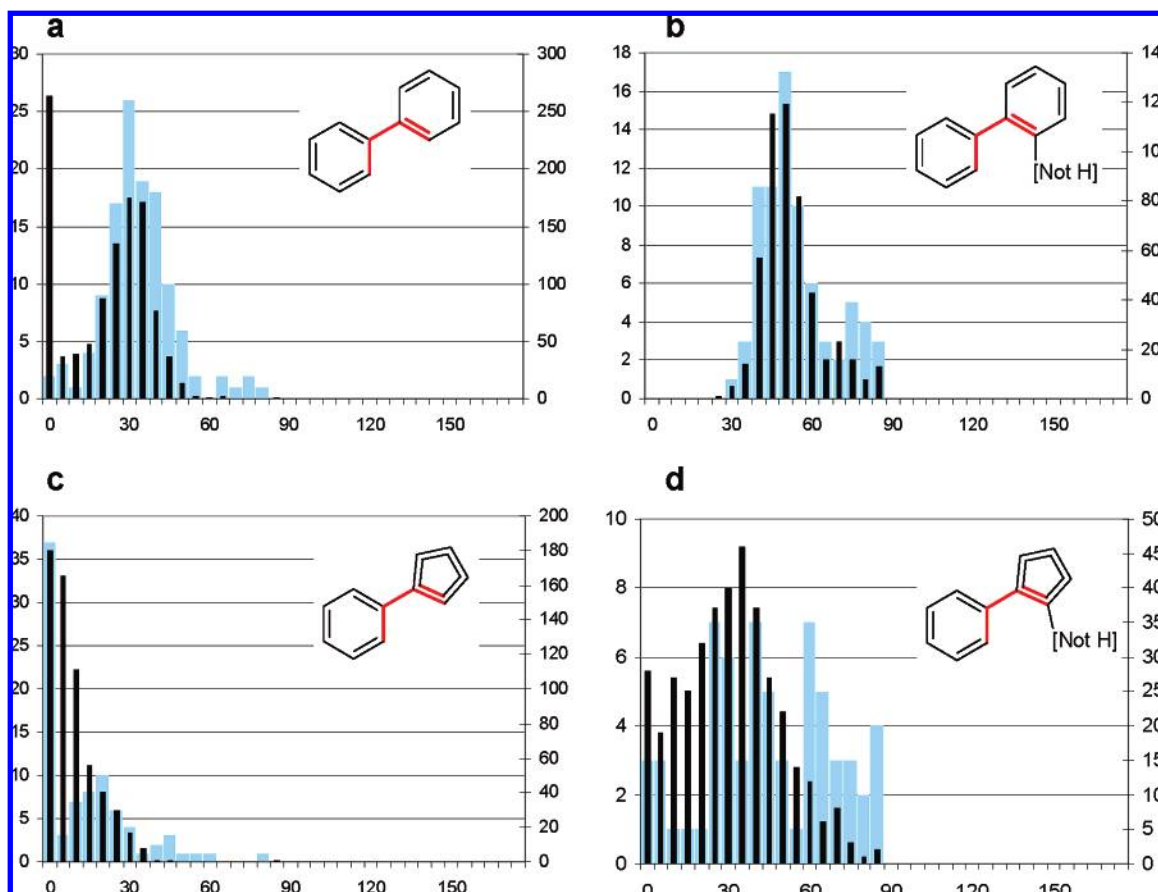


Figure 17. Torsion histograms for common biaryl ring systems. The histograms reflect the angle between the ring planes. In each case, the PDB histogram (light blue – y-axis to the left) is overlaid with the corresponding histogram derived from CSD ligands (black – y-axis to the right): (a) biphenyl systems lacking *ortho* substituents, (b) biphenyl derivatives with one *ortho*-substituent, (c) results of a composite query for various phenyl-substituted planar five-membered heterocycles lacking substituents *ortho* to the biaryl axis, and (d) query analogous to (c) with one *ortho*-substituent on the five-membered ring.

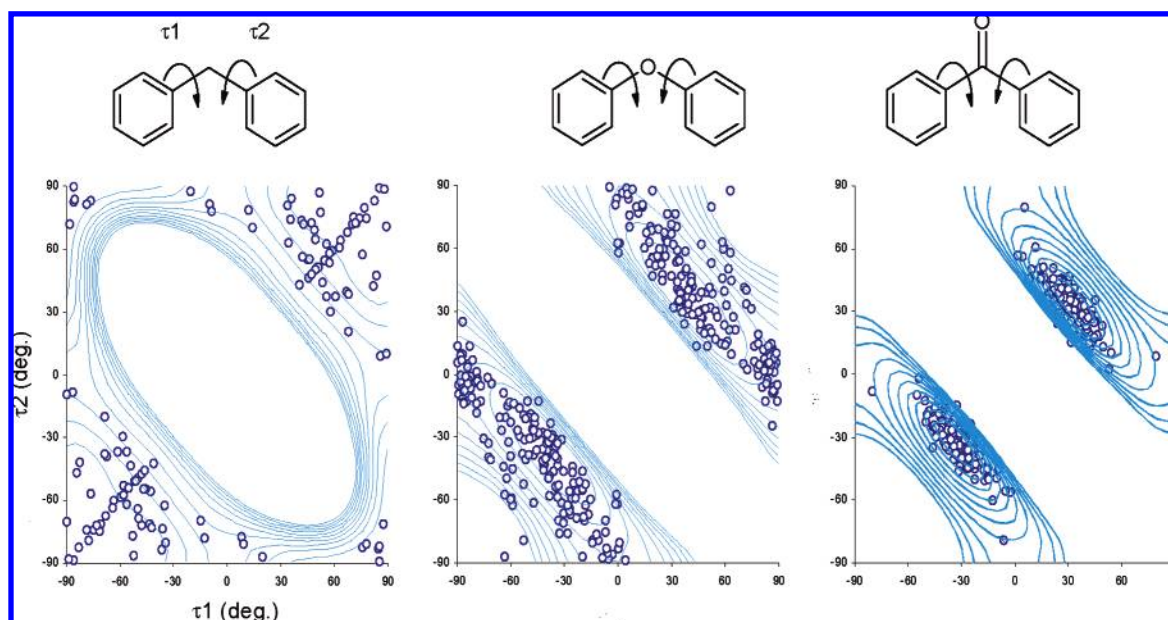


Figure 18. Torsion scatterplots for three different Ar–X–Ar systems. For each CSD structure, the pair of torsion angles with absolute values $\leq 90^\circ$ was chosen. To arrive at a better visual impression of the torsional preferences, both the original value pair and the value pair multiplied by -1 were plotted, as if the mirror image of the structure existed as a separate entry. As a consequence, the full data set is represented twice in each plot, and the upper right and lower left diagonal of the plots are mirror images of each other.

on the diagonal of the plot, indicating that a deviation from the 90° orientation on one side is generally accompanied by the same change in the other dihedral angle. This can be

interpreted as a result of avoiding steric strain caused by 1,5-repulsion of two CH units. In addition to this correlated movement, structures with one aryl system in plane and the

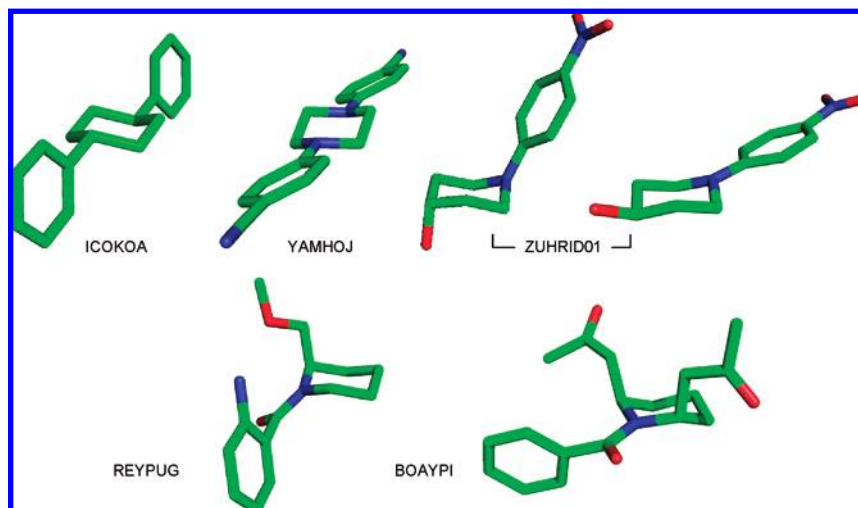


Figure 19. Upper row: CSD structures visualizing the effects of aryl substitution at cyclohexane and piperidine-type heterocycles. Bottom row: Two CSD structures illustrating the necessity of *ortho* substituents to be in axial positions in acylated piperidines and related heterocycles.

other orthogonal to its respective substituents are also observed. It can be concluded that the Ar-CH₂-Ar system is quite flexible, but structures where both rings are close to a coplanar arrangement are strongly avoided.

In contrast, the bisaryl ether and benzophenone moieties favor arrangements in which both aryl units achieve a maximum of planarity with the respective bridging group. Again, this is to be expected based on the univariate distributions of alkoxy and acyl substituents at aryl rings. As a consequence, there is now an inverse correlation between the two dihedral angles: The more planar one aryl ring is to the bridging group, the more the other one has to rotate out of plane to avoid steric strain. The benzophenone system is the most rigid of the three systems studied. Its structures and interconversion mechanisms have been explored in detail.^{86,87} Similar preferences are found for diphenylamine derivatives.

It is instructive to analyze the overlap between the torsion angle distributions of the three systems shown in Figure 18. The diphenylmethane and bisaryl ether scatterplots do overlap to a certain degree and can thus be regarded as isosteres under certain circumstances. There are very few low-energy regions common to diphenylmethane and benzophenone systems. The latter populates an area where the conformational energy of diphenylmethanes is rising strongly (dense contour lines). Thus, these two systems cannot be regarded as isosteres.

10. SIX-MEMBERED RING SYSTEMS

The basic properties of six-membered rings are textbook knowledge and will not be repeated here. Instead, we would like to focus on specific properties of N-substituted piperidine-type systems.

Structurally, one is inclined to think of *N*-phenylpiperidine as an analog of phenylcyclohexane, but there are considerable conformational differences. In the preferred conformation of phenylcyclohexane, the equatorial phenyl ring is situated on the C₃ plane of the cyclohexane chair, e.g., as in structure ICOKOA (Figure 19). The large majority of the CSD structures are within $\pm 20^\circ$ of this ideal aryl-cyclohexyl rotamer. In *N*-phenylpiperidines, the phenyl ring is generally in plane with either one of the neighboring N-C bonds of

the piperidine ring. Essentially, the aryl ring is almost in plane with the average ring plane of the piperidine ring, as in the structure YAMHOJ (Figure 19). In this fashion, a maximum degree of conjugation between the N lone pair and the phenyl ring is achieved.

Aniline nitrogen atoms are typically slightly pyramidal, the degree of pyramidalization depending strongly both on the electronic properties of the aryl ring and on the molecular environment. The more electron deficient the aryl ring, the greater is the tendency for the nitrogen atom to move toward sp² hybridization and planarity. This has important consequences for the conformational behavior of these structures. The more planar the nitrogen, the smaller are both the structural and the energy difference between a pseudoequatorial and a pseudoaxial position of the aryl ring. For 4-(4-hydroxypiperidinyl)nitrobenzene, it has been shown that the bis-equatorial and bis-axial conformation are so close in energy that they occur in the same polymorph⁸⁸ (structures ZUHRID01 in Figure 19).

Figure 20 shows a plot combining data from three CSD searches. The torsion angle between the aryl substituent and the ring is plotted against the distance between N and the plane of its substituent atoms. The upper and lower line of points represents substituents in an equatorial or axial conformation, respectively. Where they meet, the piperidine N is essentially planar and the two positions can no longer be differentiated. It can be seen that phenylpiperidines typically adopt an equatorial conformation (open circles), whereas the introduction of one (gray circles) and in particular two (black circles) *ortho* nitrogen atoms leads to an increase of sp² hybridization at the piperidine N atom and an increased relative occurrence of axial conformers, in agreement with the increased electron deficiency of the aryl ring.

These findings clearly show that arylpiperidines and related structures are different from, and more flexible than, their cyclohexane counterparts. The smaller energy difference between axial and equatorial conformations certainly allows them to adapt more freely to different protein environments. Their successful use as “privileged substructures” in particular for GPCR ligands may, in addition to their good synthetic accessibility, be founded in this general plasticity,

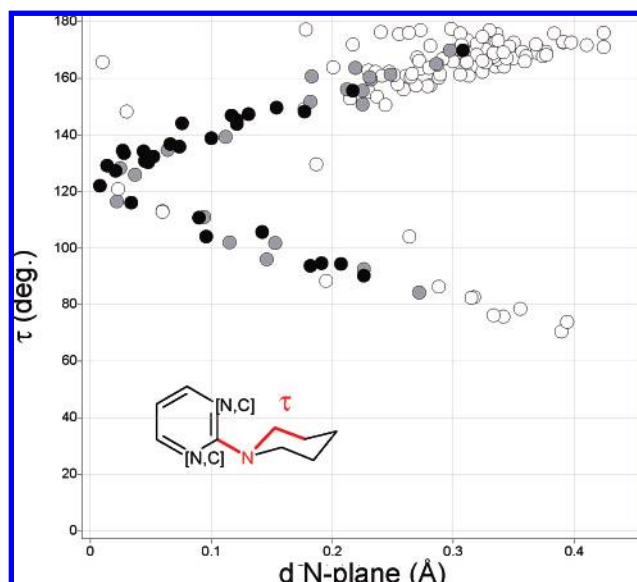


Figure 20. Analysis of *N*-phenylpiperidines and analogous heterocycles in the CSD. The torsion angle marked red in the inserted scheme is plotted against the distance between N and the plane formed by its three substituents. Open circles: phenylpiperidines, gray circles: pyridylpiperidines, black circles: pyrimidylpiperidines. The stronger the conjugation between the aniline N and the aryl ring, the less difference exists between the axial and the equatorial position in the piperidine ring.

as has been claimed for biaryl systems.⁸⁹ On a more general note, the importance of “N hybridization as a conformational variable”⁹⁰ cannot be overstated. It should be kept in mind that standard molecular mechanics force fields tend to exaggerate either the planarity or the pyramidity of nitrogen, and they cannot model the influence of substituents and the environment. Indeed, the MMFF force field exists in two versions, one of which tends to generate planar structures of conjugated nitrogen.⁹¹ One of the motivations to generate this separate version of MMFF was to appease

chemists who are used to working with experimental structures where small nonplanarity effects are not visible due to thermal averaging. We have seen above in section 7 that this attitude can lead to a dangerous self-consistent cycle where we end up believing our models more than reality.

Nitrogen atoms in *N*-acylated piperidines are typically completely planar. In Figure 20 they would be situated at the left-hand edge of the plot; there is no difference between the axial and the equatorial position at the nitrogen atom. The acyl substituent effectively blocks the equatorial positions neighboring the nitrogen atom. As a consequence, these substituents invariably occupy the axial position. This is illustrated by REYPUG with one axial substituent and BOAYPI, which even has two axial substituents in a 1,3-relationship (Figure 19). There are even structures adopting twist-boat conformations as a consequence of the steric strain imparted by the acylation (e.g., KADHAY, structure not shown). Several PDB examples illustrate that acylated piperidines or related structures have effectively been used in drug design. Examples are the thrombin complex structures 1etr and 1kli in the PDB (structures not shown).

Note that this is a third ring saturated six-membered ring system preferring axial substitutions. Before, we have encountered cyclic sulfonamides and systems governed by the anomeric effect. In addition, of course, there is the “classical” approach of driving the conformational equilibrium to one side utilizing the relative propensity of substituents of different size to adopt equatorial positions.

11. SEVEN-MEMBERED RINGS

Structures of saturated medium size rings are more difficult to analyze and visualize than those of the smaller rings, because the lowest-energy conformations undergo extensive pseudorotation, placing substituents in a number of alternative positions. Cycloheptane^{92–94} and cyclooctane^{95,96} crystal structures have been studied taking into account these

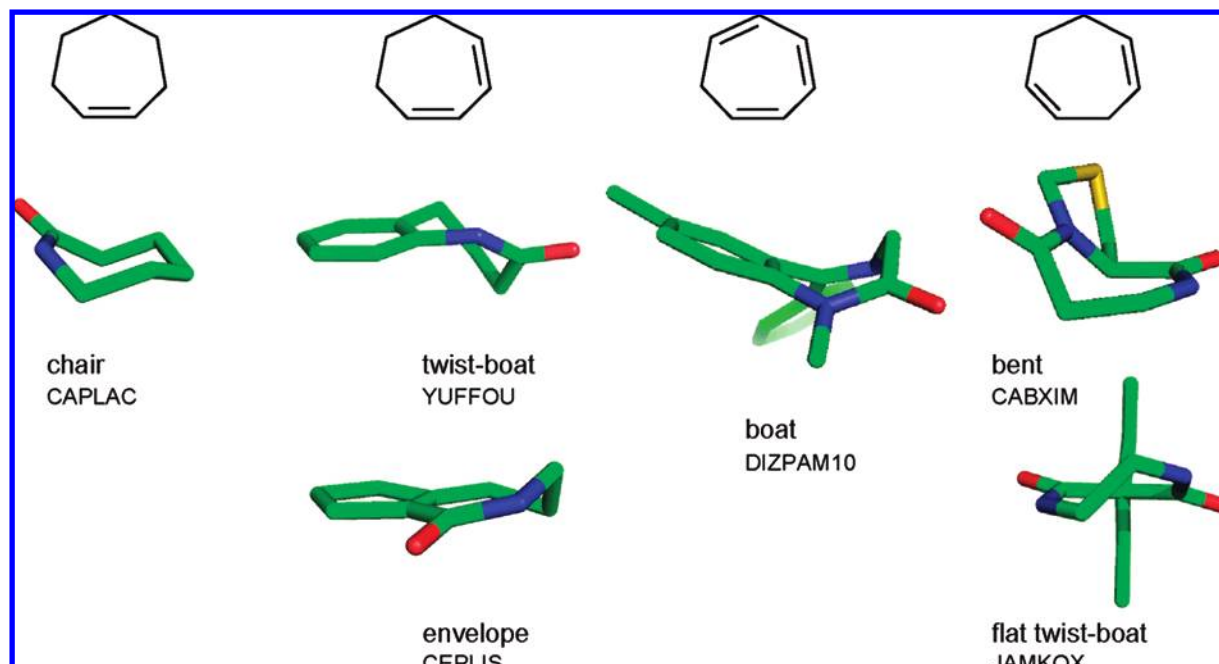


Figure 21. Representative examples of seven-membered ring systems and their preferred conformations. The key observation is that the preferences of the four possible unsaturated cycloheptane congeners also hold for analogs with double bond equivalents such as annelated rings and amide groups.

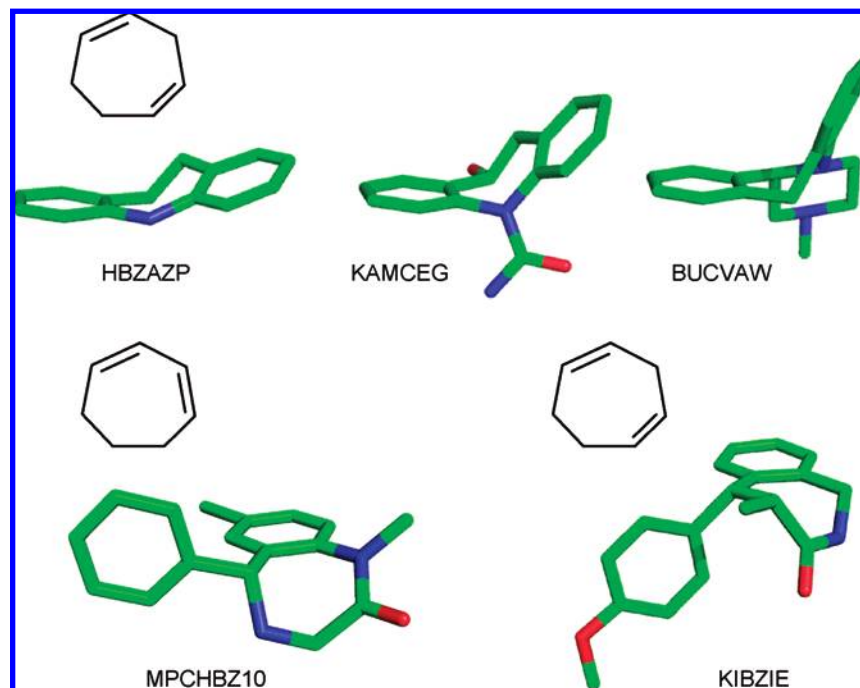


Figure 22. CSD structures illustrating the effects of the nature of bridging groups between two double bond equivalents (top row) and the shift of double bond equivalents to different positions (bottom row).

pseudorotation effects. Partially unsaturated cycloheptane derivatives are more rigid. They typically adopt only one or two basic conformations and, depending on the substitution pattern, their mirror image counterparts. Allen et al. performed systematic analyses of exocyclic⁹⁷ and endocyclic⁹⁸ unsaturated cycloheptane derivatives. Their formal approach of symmetry-modified conformational mapping and clustering has led to a set of preferred conformers for cycloheptene, cyclohepta-1,3-diene, cyclohepta-1,4-diene, and cyclohepta-1,3,5-triene. During our analysis of CSD structures, we noted that the same sets of conformations are also found for systems where double bonds are replaced by “double bond equivalents” in a broader sense; for example fused aromatic rings, amide or imine bonds, and even *cis*-fused saturated three- to five-membered rings. This makes the analysis much more relevant for medicinal chemistry, where partially unsaturated heterocyclic rings dominate, prominent examples being the diazepams and tricyclic antidepressants. Due to the complexity and heterogeneity of the available structural data, we will not attempt an exhaustive statistical analysis but rather discuss the most predominant conformation families which are likely to be relevant for these systems.

Examples of these typical conformations are displayed in Figure 21. The lowest energy conformation of cycloheptene is a chair, as was early on deduced from simple molecular mechanics consideration,⁹⁹ and this is also true for related structures such as caprolactam (CAPLAC). An alternative twist-boat conformation is rarely observed and with sterically demanding substituents only (PUZBER, GERPOI, not shown). The energy difference between the chair and the twist forms of unsubstituted cycloheptene is calculated to be 1.0 kcal/mol at the MP2/cc-pVDZ level of theory.

Systems of the cyclohepta-1,3-diene type typically occur in a twist-boat conformation (YUFFOU). For cyclohepta-1,3-diene itself, this is a C_2 -symmetric conformer.^{100,101} The alternative is the C_s -symmetric envelope conformation, where

one corner of the ring is bent away from a plane formed by the other six ring atoms. As an example, the structure CEPLIS comes very close to this conformer. The key differentiator between the two types of conformation is the torsion angle between the two double bond equivalents, which approaches zero in the envelope conformation and thus leads to an almost planar ring. Ab initio calculations at the MP2/cc-pVDZ level indicate that the two conformers of cyclohepta-1,3-diene are energetically equivalent.

Introduction of a further double bond equivalent leads to structures which exclusively adopt boat conformations, as has long been known for cyclohepta-1,3,5-triene itself.¹⁰² This is also found for diazepam (DIZPAM10, Figure 21) and many related structures.

A completely different set of conformations is found for structures of the cyclohepta-1,4-diene type. The majority of the CSD structures occur in a “bent” form, defined by two planes of five and four coplanar atoms (CABXIM). In these structures, there is a considerable twist between the two double bond equivalents. Alternatively, and depending on the nature and steric requirements of the atom linking the two double bonds, an almost coplanar arrangement of the two double bonds is found. In such cases, a flattened twist-boat conformation is formed (JAMKOX, C_2 symmetric both in this example and in cyclohepta-1,4-diene). The influence of the linking atom is quite strong, and there is a continuum of conformations between the “twisted” and “bent” extremes, as in the tricyclic structures in Figure 22: The aniline nitrogen in HBZAZP enforces conjugation with the two phenyl rings, leading to a planar twisted structure. At the other end of the scale, the structure BUCVAW adopts the bent conformation with a significantly more acute angle between the two aryl rings. The urea derivative KAMCEG lies just between these two cases. MP2/cc-pVDZ calculations on unsubstituted cyclohepta-1,4-diene indicate a somewhat higher energy of 0.9 kcal/mol for the “twisted” conformer.

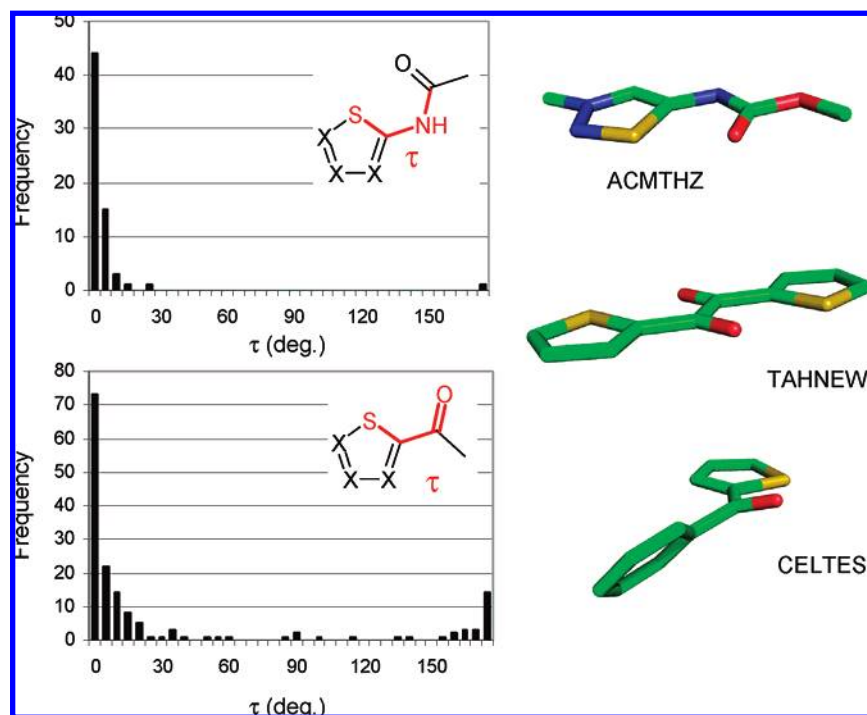


Figure 23. Torsion angle histograms and CSD examples illustrating the attractive interaction between carbonyl oxygen and sulfur in thiophene and related heterocycles.

It should be kept in mind that alternative conformations of ring systems affect not only the shape of the ring itself but also the projection of substituents relative to the ring. Figure 22 gives an example. Structure MPCHBZ10 crystallizes in the standard twist-boat conformation of cyclohepta-1,3-diene systems. KIBZIE is an analog with the amide bond moved one atom further away from the fused phenyl ring, and it adopts the typical bent conformation of cyclohepta-1,4-diene. Although the two phenyl substituents have the same absolute positions at the bicycles, they point into different directions because of the alternative local rotameric states of the ring bonds. Taking into account the key conformations and their approximate substituent positions should provide a good basis for many medicinal chemistry programs (see ref 103 for an example).

12. PREFERENCES ARISING FROM SULFUR–OXYGEN CONTACTS

Although counterintuitive at first glance, interactions between carbonyl oxygen atoms and divalent sulfur are in fact attractive: Sulfur is a large and highly polarizable atom that can favorably interact with additional electron density. This interaction is an example of closed shell interactions, which are quite well understood and have been exhaustively reviewed in particular for heavier elements.¹⁰⁴ The CSD contains a significant number (>200) of such contacts below vdW distance (<3.3 Å). For comparison, interactions below vdW distance are observed only in three cases between carbonyl C atoms and sulfur. A recent study in the PDB has revealed a clear orientation preference for thioethers and disulfide bridges relative to oxygen.¹⁰⁵ The intramolecular consequences of these interactions were highlighted by Burling et al. first on thiazole and selenazole nucleosides¹⁰⁶ and later analyzed in more detail by mining the CSD for OCCS fragments.¹⁰⁷ More recently, conformations of a series of (acylamino)thiadiazolines has been discussed by Nagao et al.¹⁰⁸

Sulfur-containing heterocycles bearing substituents with a C=O motif display a strong conformational preference for the rotamer that brings the S and the carbonyl oxygen into a *syn* relationship. Torsion histograms and selected structures of such systems are depicted in Figure 23. Acylated amino thiophenes and related structures have a particularly strong tendency for the *syn* rotamer (e.g., ACMTHZ). In fact, the only acylamino thiophene derivative in the CSD that displays an *anti* rotamer is a structure where the S–O contact competes for an intramolecular hydrogen bond (WOYQOQ, not shown). There seems to be a general rule that a 1,5-contact between S and O is more favorable than a 1,4-contact. TAHNEW is one example that has a choice between the two types of interaction, and it crystallizes in a conformation with two 1,5-contacts. Also, acylated thiophenes (e.g., CELTES) deviate more often from planarity than acylamino thiophenes. These findings are nicely confirmed by ligand structures in the PDB. Examples of ligands with 1,5 S–O contacts include an adenosine deaminase complex (1wxy), a methionine aminopeptidase complex (2evo), and a stromelysin structure (3usn). There are three Factor Xa complexes that could form 1,4-contacts, but only two of them do (1mq5, 1mq6). In the third structure (2d1j), a significant deviation from planarity is observed. The attractive nature of sulfur–oxygen contacts has also been reported in a different class of Factor Xa inhibitors,⁶¹ where it was shown that thiazole sulfonamides preferably adopt a rotameric position where one S=O bond is oriented *syn* to the ring sulfur. This tendency is nicely confirmed by more than 20 CSD entries.

13. CONCLUSIONS

We have shown that the CSD is a relevant and rich source of information on the conformational preferences of druglike organic molecules. It provides higher quality structural information and greater chemical diversity than the PDB.

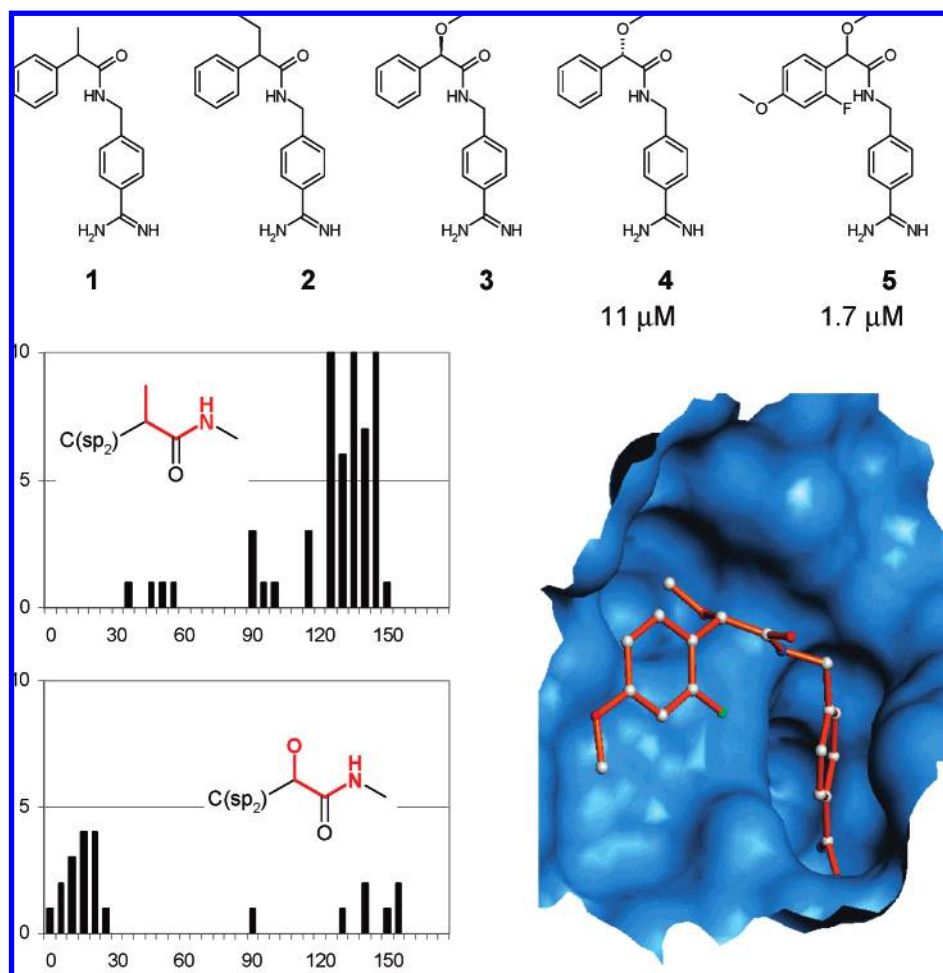


Figure 24. Case study illustrating the use of the CSD in the early phase of identifying a novel TF/F.VIIa inhibitor class. Compounds **1** and **2**, investigated initially, prefer conformations that do not fit the active site and were found to be inactive as inhibitors. It was found that their mandelic acid analogs (**4**, **5**) have different preferred conformations that do fit the binding site. These were pursued as a lead series.

Data mining of the CSD through the analysis of univariate torsion angle distributions often indicates clear trends leading to valuable hints for the design of compounds with specific shapes. However, univariate analyses should be treated with caution, and as a rule one should search for pairs of mutually dependent conformation variables. These have been exemplified in several cases, e.g., two consecutive dihedral angles as in the aryl-X-aryl systems or N pyramidal and torsion angles as in the case of sulfonamides. Furthermore, a comparison of PDB and CSD statistics has pointed out some of the biases intrinsic to ligand conformations in the PDB which are consistent with protein-induced conformational perturbations.

The main focus of structure-guided drug design is usually the optimization of interactions between a receptor and the bound ligand. Clearly, the minimization of ligand strain in the bioactive, receptor-bound conformation is equally important, as the introduction of ligand strain in the bound conformation is directly translated into reduced binding free energy. We have shown several examples from different protein targets where the knowledge about preferred ligand conformations was crucial to rationalize the observed structure–activity relationship. To improve binding affinities in molecular drug design two guidelines for the use of CSD histogram and scatterplots can be followed. First, torsional angles of ligands in the modeled receptor binding mode

should obviously fall within the distributions of the CSD. Second, more efficient binding can be expected if such a match occurs in a more narrow or single-peak torsional angle distribution, which is a consequence of a reduced ligand entropy loss upon complex formation. As an example, molecular replacement of a 2-alkoxyphenyl by a 2-alkoxy-pyridine fragment in a ligand basically abolishes the *anti* orientation (Figure 14) and, if the *syn* orientation is preferred in the receptor, leads to an additional gain in binding free energy due to reduced ligand entropy of approximately $-RT \ln 2$ (≈ 0.4 kcal/mol).

In this review, we neither attempted to comprehensively cover all substructures relevant for medicinal chemistry nor to exhaustively investigate the properties of each of the discussed substructures. Instead, we have highlighted some of those cases where, according to our own experience, a full understanding of the conformational properties has not yet become common knowledge. In particular, we hope to have covered several typical structural elements in sufficient detail to be useful for everyday medicinal chemistry work, where the data presented here might trigger more detailed investigations to address specific questions. It is our explicit wish to encourage our readers to continue exploring the universe of conformations on their own. All the analyses presented here are only the beginning of more detailed and thorough investigations. And so we would

like to conclude here with a final case study that illustrates once more the usefulness of the CSD for conformation analysis.

As part of a program to discover novel tissue factor/factor VIIa inhibitors, a team at Roche Basel investigated simple benzamide structures. Based on the well-known binding mode of the benzamidine moiety to the S1 pocket,¹⁰⁹ the team reasoned that the amide bond would form a hydrogen bond to Ser 214 at the “rim” of the S1 pocket, as observed with many thrombin inhibitors. The phenyl ring in structures like **1** or **2** (Figure 24) should bend off toward the S3 pocket, while the small alkyl group in the benzylic position should point into the S2 pocket specific to the F.VIIa active site.¹¹⁰ Neither **1** nor **2**, however, showed any activity within the detection limit of about 70 μ M in the chromogenic substrate assay. More detailed investigations of the anticipated binding mode showed that the required nearly coplanar arrangement between the alkyl group and amide nitrogen was not favorable. CSD searches for the corresponding fragment (Figure 24) show a strong preference for torsion angles around 120–150°, even if the overall occurrence is low. The corresponding mandelic acid fragment, where the alkyl group is replaced by an alkoxy group, does prefer the desired rotamer (calculations and X-ray studies on a number of mandelic acid derivatives are also discussed in ref 111 with the same conclusions). Indeed, it was found that the *S* enantiomer of the methoxy derivative **4** was an 11 μ M inhibitor of TF/F.VIIa. The corresponding *R* enantiomer **3** was inactive. These findings provided the entry into a new chemical class, which was subsequently further explored. With the more potent derivative **5**, the validity of the modeled binding mode could be demonstrated (Figure 24). This example shows that very specific substructure searches in the CSD can help to arrive at a better understanding of conformational space.

ACKNOWLEDGMENT

This work would not have been possible without the support from many colleagues in the medicinal chemistry, molecular structure, and molecular design departments at all Roche research sites. In particular, the authors thank Prof. Klaus Müller (Roche Basel) and Prof. François Diederich (ETH Zürich) for many stimulating discussions and Dr. Ulrike Obst for the F.VIIa example in the Conclusions section.

Supporting Information Available: Expanded Figures 3 and 6 with PDB histograms added. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- Leach, A. R. A Survey of Methods for Searching the Conformational Space of Small and Medium-sized Molecules. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, pp 1–55.
- Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, B58, 380–388.
- Dunitz, J. D.; Bürgi, H.-B. *Structure Correlation*; VCH: Weinheim, 1994.
- Lauri, G.; Bartlett, P. A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput.-Aided Mol. Des.* **1994**, 8, 51–66.
- Maass, P.; Schulz-Gasch, T.; Stahl, M.; Rarey, M. Recore: A Fast and Versatile Method for Scaffold Hopping Based on Small Molecule Crystal Structure Conformations. *J. Chem. Inf. Model.* **2007**, 47, 390–399.
- Allen, F. H.; Motherwell, W. D. S. Applications of the Cambridge Structural Database in organic chemistry and crystal chemistry. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, B58, 407–422.
- Taylor, R. Life-science applications of the Cambridge Structural Database. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, D58, 879–888.
- Allen, F. H.; Taylor, R. Research applications of the Cambridge Structural Database (CSD). *Chem. Soc. Rev.* **2004**, 33, 463–475.
- CCDC Website database. http://www.ccdc.cam.ac.uk/free_services/webcite/ (accessed July 2, 2007).
- Bernstein, F. C.; Koetzle, T. E.; Williams, G. J. B.; Meyer, J., E. F.; Brice, M. D. et al. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, 112, 535–542.
- Protein Data Bank. <http://www.rcsb.org/pdb/> (accessed July 2, 2007).
- Brock, C. P.; Minton, R. P. Systematic Effects of Crystal-Packing Forces: Biphenyl Fragments with H Atoms in All Four Ortho Positions. *J. Am. Chem. Soc.* **1989**, 111, 4586–4593.
- 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.
- Boehm, H.-J.; Klebe, G. What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs? *Angew. Chem., Int. Ed.* **1996**, 35, 2588–2614.
- Klebe, G.; Mietzner, T. A fast and efficient method to generate biologically relevant conformations. *J. Comput.-Aided Mol. Des.* **1994**, 8, 583–606.
- Omega, Version 1.8.1; Open Eye: Santa Fe, NM. <http://www.eyesopen.com/> (accessed July 2, 2007).
- Sadowski, J.; Bostroem, J. MIMUBA Revisited: Torsion Angle Rules for Conformer Generation Derived from X-ray Structures. *J. Chem. Inf. Model.* **2006**, 46, 2305–2309.
- SMARTS - A Language Describing Molecular Patterns. <http://www.daylight.com> (accessed July 2, 2007).
- Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F. et al. New software for searching the Cambridge structural database and visualizing crystal structures. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, B58, 389–397.
- Microsoft Excel 2003. www.microsoft.com (accessed July 2, 2007).
- Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. et al. *Gaussian 98, Revision A.6*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- Clark, T.; Alex, A.; Beck, B.; Burckhardt, F.; Chandrasekhar, J. et al. *VAMP 8.0*; Erlangen, Germany, 2001.
- Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 244–255.
- Desert Scientific Software. <http://www.desertsci.com/> (accessed July 2, 2007).
- PreQuest. http://www.ccdc.cam.ac.uk/products/csd_system/prequest/ (accessed July 2, 2007).
- Blow, D. M. Rearrangement of Cruickshank's formulae for the diffraction-component precision index. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, D58, 792–797.
- Prous Science, S. A. Prous Science Integrity: Barcelona, Spain, 2007.
- Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, 39, 2287–2893.
- Clark, D. E.; Pickett, S. D. Computational methods for the prediction of “drug-likeness”. *Drug Discovery Today* **2000**, 5, 49–58.
- Brüistle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T. et al. Descriptors, physical properties, and drug-likeness. *J. Med. Chem.* **2002**, 45, 3345–3355.
- Structural elements containing sugar rings were deliberately omitted due to their limited importance in current drug design.
- Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; John Wiley & Sons, Inc.: New York, 1994.
- Wiberg, K. B.; Murcko, M. A. Rotational Barriers. 2. Energies of Alkane Rotamers. An Examination of Gauche Interactions. *J. Am. Chem. Soc.* **1988**, 110, 8029–8038.
- Hoffmann, R. W.; Stahl, M.; Schopfer, U.; Frenking, G. Conformation design of hydrocarbon backbones. *Chem. Eur. J.* **1998**, 4, 559–566.
- Anderson, J. E. Why Do so Few Simple Acyclic Acetals Adopt the Classic Anomeric Conformation? The Eclipsed Anomeric Conformation for Acetals. An Analysis of Crystal Structures, Molecular Mechanics Calculations, and NMR Measurements. *J. Org. Chem.* **2000**, 65, 748–754.
- Juaristi, E.; Cuevas, G. *The Anomeric Effect*; CRC Press, Inc.: Boca Raton, FL, 1995.
- Fuchs, B.; Schleifer, L.; Tartakovsky, E. Probing the Anomeric Effect: The Structural Criterion. *New J. Chem.* **1984**, 8, 275–278.
- Cossé-Barbi, A.; Dubois, J.-E. Anomeric Orbital and Steric Control in Static Conformations and Systems Dynamics: Rotations of Methoxy

- Groups in 2,2-Dimethoxypropane and Similar Crystallographic CO-COC Fragments. *J. Am. Chem. Soc.* **1987**, *109*, 1503–1511.
- (39) Senderowitz, H.; Aped, P.; Fuchs, B. Probing the Anomeric Effect in O-C-N Systems: Theory vs. Experiment: MO-ab initio Calculations and a Structural-Statistical Analysis. *Helv. Chim. Acta* **1990**, *73*, 2113–2128.
- (40) Senderowitz, H.; Aped, P.; Fuchs, B. Computation of N-C-N Systems: Theory vs. Experiment. *Tetrahedron* **1992**, *48*, 1131–1144.
- (41) Salzner, U.; Schleyer, P. v. R. Generalized anomeric effects and hyperconjugation in CH₂(OH)₂, CH₂(SH)₂, CH₂(SeH)₂, and CH₂(TeH)₂. *J. Am. Chem. Soc.* **1993**, *115*, 10231–10236.
- (42) Wolfe, S. The Gauche Effect. Some Stereochemical Consequences of Adjacent Electron Pairs and Polar Bonds. *Acc. Chem. Res.* **1972**, *5*, 102–111.
- (43) Kirby, A. J. *The Anomeric Effect and Related Stereochemical Effects*; Springer-Verlag: Berlin, Germany, 1983.
- (44) Amos, R. D.; Handy, N. C.; Jones, P. G.; Kirby, A. J.; Parker, J. K. et al. Bond Length and Reactivity: the gauche Effect. A Combined Crystallographic and Theoretical Investigation of the Effects of β -Substituents on C-OX Bond Length. *J. Chem. Soc., Perkin Trans. 2* **1992**, 549–558.
- (45) Partington, P.; Feeney, J.; Burgen, A. S. V. The Conformation of Acetylcholine and Related Compounds in Aqueous Solution as Studied by Nuclear Magnetic Resonance Spectroscopy. *Mol. Pharmacol.* **1972**, *8*, 269–277.
- (46) Behling, R. W.; Yamane, T.; Navon, G.; Jelinski, L. W. Conformation of acetylcholin bound to the nicotinic acetylcholine receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6721–6725.
- (47) Bourne, Y.; Radic, Z.; Sulzenbacher, G.; Kim, E.; Taylor, P. et al. Substrate and Product Trafficking through the Active Center Gorge of Acetylcholinesterase Analyzed by Crystallography and Equilibrium Binding. *J. Biol. Chem.* **2006**, *281*, 29256–29267.
- (48) Johnson, F.; Malhotra, S. K. Steric Interference in Allylic and Pseudo-Allylic Systems. I. Two Stereochemical Theorems. *J. Am. Chem. Soc.* **1965**, *87*, 5492–5493.
- (49) Johnson, F. Allylic Strain in Six-membered Rings. *Chem. Rev.* **1968**, *68*, 375–413.
- (50) Hoffmann, R. W. Allylic 1,3-strain as a controlling factor in stereo-selective transformations. *Chem. Rev.* **1989**, *89*, 1841–1860.
- (51) Sintchak, M. D.; Fleming, M. A.; Futer, O.; Raybuck, S. A.; Chambers, S. P. et al. Structure and Mechanism of Inosine Monophosphate Dehydrogenase in Complex with the Immunosuppressant Mycophenolic Acid. *Cell* **1996**, *85*, 921–930.
- (52) Chakrabarti, P.; Dunitz, J. D. Structural Characteristics of the Carboxylic Amide Group. *Helv. Chim. Acta* **1982**, *65*, 1555–1562.
- (53) Schweizer, W. B.; Dunitz, J. D. Structural Characteristics of the Carboxylic Ester Group. *Helv. Chim. Acta* **1982**, *65*, 1547–1554.
- (54) Li, G.; Hasvold, L. A.; Tao, Z.-F.; Wang, G. T.; Gwaltney, II, S. L. et al. Synthesis and biological evaluation of 1-(2,4,5-trisubstituted phenyl)-3-(5-cyanopyrazin-2-yl)ureas as potent Chk1 kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2293–2298.
- (55) Tao, Z.-F.; Wang, L.; Stewart, K. D.; Chen, Z.; Gu, W. et al. Structure-Based Design, Synthesis, and Biological Evaluation of Potent and Selective Macrocyclic Checkpoint Kinase 1 Inhibitors. *J. Med. Chem.* **2007**, *50*, 1514–1527.
- (56) Sandrone, G.; Dixon, D. A.; Hay, B. P. C(sp²)-C(sp³) Rotational Barriers in Simple Amides: H₂N-C(=O)-R (R = Methyl, Ethyl, i-Propyl, tert-Butyl). *J. Phys. Chem. A* **1999**, *103*, 893–902.
- (57) Banks, J. W.; Batsanov, A. S.; Howard, J. A. K.; O'Hagan, D.; Rzepa, H. S. et al. The preferred conformation of alpha-fluoroamides. *J. Chem. Soc., Perkin Trans. 2* **1999**, 2409–2411.
- (58) Hao, M.-H.; Haq, O.; Muegge, I. Torsion Angle Preference and Energetics of Small-Molecule Ligands Bound to Proteins. *J. Chem. Inf. Model.* **2007**, *47*, 2242–2252.
- (59) Beddoes, R. L.; Dalton, L.; Joule, J. A.; Mills, O. S.; Street, J. D. et al. The Geometry at Nitrogen in N-Phenylsulphonyl-pyrroles and -indoles. The Geometry of Sulphonamides. *J. Chem. Soc., Perkin Trans. 2* **1986**, 787–797.
- (60) Menziani, M. C.; Cocchi, M.; De Benedetti, P. G. Electronic and electrostatic aspects of carbonic anhydrase inhibition by sulphonamides. *J. Mol. Struct.* **1992**, *256*, 217–229.
- (61) Senger, S.; Convery, M. A.; Chan, C.; Watson, N. S. Arylsulfonamides: A study of the relationship between activity and conformational preferences for a series of factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5731–5735.
- (62) Senger, S.; Chan, C.; Convery, M. A.; Hubbard, J. A.; Shah, G. P. et al. Sulfonamide-related conformational effects and their importance in structure-based design. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2931–2934.
- (63) Jorgensen, W. L.; Salem, L. *The organic chemist's book of orbitals*; Academic Press: New York and London, 1973.
- (64) Bombicz, P.; Czugler, M.; Kalman, A.; Kapovits, I. A Database Study of the Bonding and Conformation of Bis-sulfonylamide/imide Moieties. *Acta Crystallogr., Sect. B: Struct. Sci.* **1996**, *B52*, 720–727.
- (65) Boer, D. R.; Kroon, J.; Cole, J. C.; Smith, B.; Verdonk, M. L. SuperStar: Comparison of CSD and PDB-based interaction fields as a basis for the prediction of protein-ligand interactions. *J. Mol. Biol.* **2001**, *313*, 275–287.
- (66) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* **1995**, *3*, 411–428.
- (67) Vieth, M.; Hirst, J. D.; Brooks, C. L., III Do active site conformations of small ligands correspond to low free-energy solution structures? *J. Comput.-Aided Mol. Des.* **1998**, *12*, 563–572.
- (68) Bostrom, J.; Norrby, P.-O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383–396.
- (69) Perola, E.; Charifson, P. A. Conformational sampling of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.
- (70) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem., Int. Ed.* **2003**, *42*, 2718–2736.
- (71) Vargas, R.; Garza, J.; Dixon, D.; Hay, B. P. C(sp²)-C(Aryl) Bond Rotation Barrier in N-Methylbenzamide. *J. Phys. Chem. A* **2001**, *105*, 774–778.
- (72) Pophristic, V.; Vemparala, S.; Ivanov, I.; Liu, Z.; Klein, M. L. et al. Controlling the shape and flexibility of arylamides: A combined ab initio, molecular dynamics, and classical molecular dynamics study. *J. Phys. Chem. B* **2006**, *110*, 3517–3526.
- (73) Vargas, R.; Garza, J.; Dixon, D.; Hay, B. P. Conformational analysis of N-benzylformamide. *J. Mol. Struct.: THEOCHEM* **2001**, *541*, 243–251.
- (74) Hummel, W.; Roszak, A.; Bürgi, H.-B. Conformational Flexibility of the Acetoxyphenyl Group Studies by Statistical Analysis of Crystal Structure Data. *Helv. Chim. Acta* **1988**, *71*, 1281–1290.
- (75) For both parts c and d of Figure 12 those PDB structures were excluded where the sulfonamide N is in close contact with a metal ion (typically zinc in carbonic anhydrase structures). In these cases, the sulfonamide is deprotonated and typically adopts a different rotamer. It can, however, not be excluded that there are deprotonated sulfonamides among the remaining structures, which might slightly affect the PDB distributions.
- (76) Tsutsuki, S.; Houjou, H.; Nagawa, Y.; Hiratani, K. High-Level ab Initio Calculations of Torsional Potential of Phenol, Anisole, and o-Hydroxyanisole: Effects of Intramolecular Hydrogen Bond. *J. Phys. Chem. A* **2000**, *104*, 1332–1336.
- (77) Tsutsuki, S.; Houjou, H.; Nagawa, Y.; Hiratani, K. The second stable conformation of the methoxy groups of o-dimethoxybenzene: stabilization of perpendicular conformation by CH–O interaction. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1271–1273.
- (78) Klocker, J.; Karpfen, A.; Wolschann, P. On the structure and torsional potential of trifluoromethoxybenzene: an ab initio and density functional study. *Chem. Phys. Lett.* **2003**, *367*, 566–575.
- (79) Hummel, W.; Huml, K.; Bürgi, H.-B. Conformational Flexibility of the Methoxyphenyl Group Studies by Statistical Analysis of Crystal Structure Data. *Helv. Chim. Acta* **1988**, *71*, 1291–1302.
- (80) Reimann, B.; Buchhold, K.; Barth, H.-D.; Brutschy, B. Anisole-(H₂O)_n (n = 1–3) complexes: An experimental and theoretical investigation of the modulation of optimal structures, binding energies, and vibrational spectra in both the ground and first excited states. *J. Chem. Phys.* **2002**, *117*, 8805–8822.
- (81) Richardson, C. M.; Williamson, D. S.; Parratt, M. J.; Borgognoni, J.; Cansfield, A. D. et al. Triazolo[1,5-a]pyrimidines as novel CDK2 inhibitors: Protein structure-guided design and SAR. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1353–1357.
- (82) Yin, H.; Hamilton, A. D. Strategies for targeting protein-protein interactions with synthetic agents. *Angew. Chem., Int. Ed.* **2005**, *44*, 4130–4163.
- (83) Desiraju, G. R.; Steiner, T. *The weak hydrogen bond in chemistry and biology*; Oxford University Press: Oxford, U.K., 1999.
- (84) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase inhibitors and the case for CH...O hydrogen bonds in protein-ligand binding. *Proteins* **2002**, *49*, 567–576.
- (85) Bettencourt-Dias, A.; Viswanathan, S.; Ruddy, K. Intermolecular Forces and Functional Group Effects in the Packing Structure of Thiophene Derivatives. *Cryst. Growth Des.* **2005**, *5*, 1477–1483.
- (86) Gough, K. M.; Wildman, T. A. Hindered Internal Rotation in Benzophenone. *J. Am. Chem. Soc.* **1990**, *112*, 9141–9144.
- (87) Rappoport, Z.; Biali, S. E.; Kaftory, M. Application of the Structural Correlation Method to Ring-Flip Processes in Benzophenones. *J. Am. Chem. Soc.* **1990**, *112*, 7742–7748.

- (88) Sharma, S.; Radhakrishnan, T. P. Modeling Polymorphism-Solvated Supramolecular Clusters Reveal the Solvent Selection of SHG Active and Inactive Dimorphs. *J. Phys. Chem. B* **2000**, *104*, 10191–10195.
- (89) Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **2000**, *43*, 3443–3447.
- (90) Andrews, P. R.; Munro, S. L. A.; Sadek, M.; Wong, M. G. The Hybridization State of Nitrogen as a Conformational Variable in Biologically Active Molecules. *J. Chem. Soc., Perkin Trans. 2* **1988**, 711–718.
- (91) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
- (92) Sauriol-Lord, F.; Grindley, T. B. A New Approach to the Conformational Analysis of Seven-Membered Rings. *J. Am. Chem. Soc.* **1981**, *103*, 936–938.
- (93) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A. Symmetry-Modified Conformational Mapping and Classification of the Medium Rings from Crystallographic Data. I. Cycloheptane. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *B49*, 910–928.
- (94) Entrena, A.; Campos, J.; Gomez, J. A.; Gallo, M. A.; Espinosa, A. A New Systematization of the Conformational Behaviour of Seven-Membered Rings. Isoclinical Anomeric and Related Orientations. *J. Org. Chem.* **1997**, *62*, 337–349.
- (95) Allen, F. H.; Howard, J. A. K.; Pitchford, N. A. Symmetry-modified Conformational Mapping and Classification of the Medium Rings from Crystallographic Data. IV. Cyclooctane and Related Eight-Membered Rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **1996**, *B52*, 882–891.
- (96) Perez, J.; Nolsoe, K.; Kessler, M.; Garcia, L.; Perez, E. et al. Bayesian methods for the conformational classification of eight-membered rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *B61*, 585–594.
- (97) 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. II. *exo*-Unsaturated and Heterocyclic Seven-Membered Rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **1994**, *B50*, 382–395.
- (98) Allen, F. H.; Garner, S. E. Symmetry-Modified Conformational Mapping and Classification of the Medium Rings from Crystallographic Data. III. *endo*-Unsaturated Seven-Membered Rings. *Acta Crystallogr., Sect. B: Struct. Sci.* **1994**, *B50*, 395–404.
- (99) Favini, G.; Buemi, G.; Raimondi, M. Molecular conformation of cyclenes. I. Cyclohexene, cycloheptene, *cis*- and *trans*-cyclooctene, *cis*- and *trans*-cyclononene. *J. Mol. Struct.* **1968**, *2*, 137–148.
- (100) Allinger, N. L.; Sprague, J. T. On the molecular structure of cycloheptadiene. *Tetrahedron* **1973**, *29*, 3811–3812.
- (101) Nevins, N.; Stewart, E. L.; Allinger, N. L.; Bowen, J. P. Ab initio and molecular mechanics calculations on the inversion of C_s to C₂ conformations of 1,3-cycloheptadiene. *J. Phys. Chem.* **1994**, *98*, 2056–2061.
- (102) Anet, F. A. L. Ring inversion in cycloheptatriene. *J. Am. Chem. Soc.* **1964**, *86*, 458–460.
- (103) Keenan, R. M.; Callahan, J. F.; Samanen, J. M.; Bondinell, W. E.; Calvo, R. R. et al. Conformational Preferences in a Benzodiazepine Series of Potent Nonpeptide Fibrinogen Receptor Antagonists. *J. Med. Chem.* **1999**, *42*, 545–559.
- (104) Pyykkö, P. Strong closed-shell interactions in inorganic chemistry. *Chem. Rev.* **1997**, *97*, 597–636.
- (105) Iwaoka, M.; Takemoto, S.; Okada, M.; Tomoda, S. Weak nonbonded S...X (X = O, N, and S) interactions in proteins. Statistical and theoretical studies. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 1611–1625.
- (106) Burling, F. T.; Goldstein, B. M. Computational Studies of Nonbonded Sulfur-Oxygen and Selenium-Oxygen Interactions in the thiazole and Selenazole Nucleosides. *J. Am. Chem. Soc.* **1992**, *114*, 2313–2320.
- (107) Burling, F. T.; Goldstein, B. M. A Database Study of Nonbonded Intramolecular Sulfur-Nucleophile Contacts. *Acta Crystallogr., Sect. B: Struct. Sci.* **1993**, *B49*, 738–744.
- (108) Nagao, Y.; Hirata, T.; Goto, S.; Sano, S.; Kakehi, A. et al. Intramolecular Nonbonded S O Interaction Recognized in (Acylimino)-thiadiazoline Derivatives as Angiotensin II Receptor antagonist and Related Compounds. *J. Am. Chem. Soc.* **1998**, *120*, 3104–3110.
- (109) Leung, D.; Abbenante, G.; Fairlie, D. P. Protease inhibitors: Current status and future prospects. *J. Med. Chem.* **2000**, *43*, 305–341.
- (110) Groebke, Zbinden, K.; Banner, D. W.; Hilpert, K.; Himber, J.; Lavé, T. et al. Dose-dependent antithrombotic activity of an orally active tissue factor/factor VIIa inhibitor without concomitant enhancement of bleeding propensity. *Bioorg. Med. Chem. Lett.* **2006**, *14*, 5357–5369.
- (111) Salas-Coronado, R.; Vasquez-Badillo, A.; Medina-Garcia, M.; Garcia-Colon, J. G.; Nöth, H. et al. Hydrogen bonds and preferred conformation of optically active amides. *J. Mol. Struct.* **2001**, *543*, 259–275.

CI7002494