Journal of Medicinal Chemistry
Article

J. Med. Chem. **2010,** *53,* 2601–2611 **2601**
DOI: 10.1021/jm100087s

Intramolecular Hydrogen Bonding in Medicinal Chemistry

Bernd Kuhn,* Peter Mohr, and Martin Stahl

Discovery Chemistry, F. Hoffmann-La Roche AG, CH-4070 Basel, Switzerland

Received January 20, 2010

The formation of intramolecular hydrogen bonds has a very pronounced effect on molecular structure and properties. We study both aspects in detail with the aim of enabling a more rational use of this class of interactions in medicinal chemistry. On the basis of exhaustive searches in crystal structure databases, we derive propensities for intramolecular hydrogen bond formation of five- to eight-membered ring systems of relevance in drug discovery. A number of motifs, several of which are clearly underutilized in drug discovery, are analyzed in more detail by comparing small molecule and protein−ligand X-ray structures. To investigate effects on physicochemical properties, sets of closely related structures with and without the ability to form intramolecular hydrogen bonds were designed, synthesized, and characterized with respect to membrane permeability, water solubility, and lipophilicity. We find that changes in these properties depend on a subtle balance between the strength of the hydrogen bond interaction, geometry of the newly formed ring system, and the relative energies of the open and closed conformations in polar and unpolar environments. A number of general guidelines for medicinal chemists emerge from this study.

Introduction

Druglike organic molecules almost invariably contain a number of functional groups capable of forming hydrogen bonds, rendering them soluble and giving them the ability to form specific interactions with their biomolecular targets. When a donor and an acceptor are in proximity on the same molecule, an equilibrium may exist between closed conformations in which an intramolecular hydrogen bond is formed, creating a temporary ring system, and open conformations in which the polar groups are exposed to solvent (Figure 1). These sets of conformations are not only structurally distinct. It is intuitively clear that the closed forms, hiding polarity from the environment, should be more lipophilic and might display a higher membrane permeability, whereas the open forms should be more water-soluble. In drug discovery, it is therefore important to recognize the potential for intramolecular hydrogen bond formation and to be aware of its consequences. Here we perform a systematic study of topologies prone to intramolecular hydrogen bond formation to enable the rational use of such interactions in medicinal chemistry. To this end, we combine data mining in crystal structure databases with the detailed analysis of measured and calculated molecular properties in illustrative model systems.

Several publications have reported beneficial effects of intramolecular hydrogen bonding on ligand−receptor binding and rationalized this with conformational restriction in which the small molecule substituents are favorably aligned with the protein pockets.[1−3] In these cases, the intramolecular hydrogen bond is crucial and removal of either hydrogen bond acceptor or donor results in a drastic loss of binding affinity. Rational design of internal hydrogen bonds for conformational preorganization was pursued in scaffold replacements for a diverse set of targets.[4−8] In addition to its effects on receptor binding, an improved property profile has been associated with molecules containing intramolecular hydrogen bonds. Increased brain penetration and pharmacological activity was observed for NK1[a] receptor antagonists and attributed to the higher apparent lipophilicity due to intramolecular hydrogen bonding.[9] A team at Takeda has reported improved oral absorption and an excellent pharmacokinetic profile for luteinizing hormone-releasing hormone receptor antagonists when an intramolecular hydrogen bond was established.[10] Studies on cyclic peptides support the notion that the ability to form internal hydrogen bonds is critical for passive membrane permeability.[11]

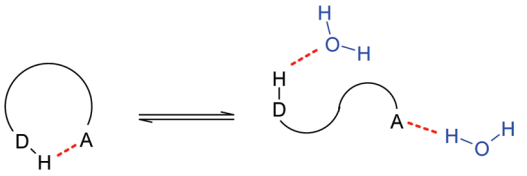


Figure 1. Molecules with hydrogen bond donor (D−H) and acceptor (A) functionalities in proximity often occur in a thermodynamic equilibrium between closed (left) and open (right) conformations.

Databases of crystal structures are rich sources of information on molecular interactions and conformations. Infantes et al. have investigated the competition between inter- and intramolecular hydrogen bonding and concluded that

*To whom correspondence should be addressed. Phone: +41 616889773. Fax: +41 616886459. E-mail: bernd.kuhn@roche.com.

[a] Abbreviations: NK1, neurokinin 1; CSD, Cambridge Structural Database; PDB, Protein Data Bank; B3LYP, Becke three-parameter, Lee, Yang, and Parr; MMFF94x, Merck molecular force field 94x; DMF, dimethylformamide; THF, tetrahydrofuran; LYSA, lyophilized solubility assay; PAMPA, parallel artificial membrane permeability assay.

© 2010 American Chemical Society

Published on Web 02/22/2010

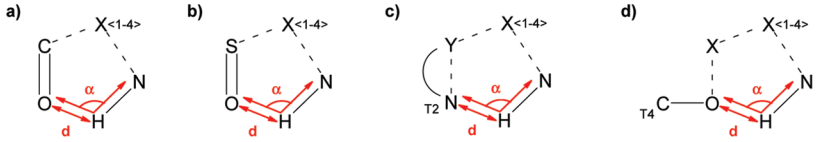


Figure 2. CSD search queries for NH hydrogen bond donor vs different acceptors: (a) C=O; (b) S=O; (c) disubstituted nitrogen acceptor in a heterocycle; (d) alkoxy group. X stands for any atom, Y for carbon or nitrogen. Subscripts denote coordination numbers. Semicircles represent cyclic bonds. Hydrogen bonds are defined by the geometric constaints $d \leq 2.35$ Å and $100° \leq \alpha \leq 180°$.

intramolecular hydrogen bonds are preferred when five- or six-membered conjugated rings are formed.[12] A pioneering statistical analysis of intramolecular hydrogen bonds in the Cambridge Structural Database (CSD)[13] was performed by Bilton et al.[14] This analysis resulted in 50 intramolecular hydrogen bond topologies and their propensity of formation within small molecule crystal structures. However, a considerable number of these motifs are not relevant to medicinal chemistry because of undesired substructure liabilities. Our search focuses exclusively on motifs of interest to pharmaceutical drug design, and by profiting from the more than 2-fold higher number of entries in the current database, we are able to identify several additional motifs. Moreover, we illustrate the frequently occurring motifs with selected crystal structures from both CSD and Protein Data Bank (PDB).[15,16]

Experimental Section

Database Searches. All statistical data described in this article were derived with the ConQuest 1.11 program. For small molecule crystal structures the CSD, version 5.30 (November 2008), was searched. Unless noted otherwise, the following general search flags were set: $R \leq 0.10$, "3D coordinates determined", "not disordered", "no ions", "no errors", "not polymeric", and "only organic". This results in a search space of 124 791 structures. Conformations of protein-bound ligands were analyzed by extracting ligands in SD format from Proasis2,[17] a curated version of the PDB, and by generating a separate database in CSD format with the program PreQuest. The ligand structures were taken from all HET entries, except metals and commonly found small ions, in the PDB as of December 8, 2008. Ligands with less than 5 or more than 100 atoms were removed, excluding in particular large peptidic ligands. A resolution limit of 3.0 Å was applied, resulting in a database of 73 158 ligands derived from 23 291 PDB entries. CSD search queries are depicted in Figure 2.

The CSD database contains a number of entries that have substructures typically not found in drugs. To focus on the chemical space relevant for medicinal chemistry, we report only those intramolecular hydrogen bond motifs that are present to a larger extent (> 5 hits) in a filtered Prous Integrity database.[18] This filtered database contains all 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", "phase I/II", "phase II", "phase II/III", "phase III", "pre-registered", "recommended approval", "registered", or "launched". Molecular weight was limited between 250 and 800. Salts and entries with elements other than H, C, N, O, F, S, P, Cl, Br, and I were removed. SD files of the resulting 4555 compounds were converted into a Conquest-searchable database using Prequest to be able to apply consistent search queries for all databases.

We generated CSD statistics for all five- to eight-membered intramolecular hydrogen bond motifs of C=O, S=O, disubstituted nitrogen acceptor in a heterocycle, and alkoxy groups as hydrogen bond acceptors with NH donors. For OH donors, additional searches were performed for five- to seven-membered rings with carbonyl and nitrogen as acceptors. Geometric cutoffs were used to classify interactions as hydrogen bonds: upper limits $d = 2.35$ Å for N−H···N/O and $d = 2.30$ Å for O−H···N/O contacts[14] were used, and the angle α at the hydrogen atom was confined to $100° \leq \alpha \leq 180°$ (Figure 2). For the PDB database, there is more ambiguity in identifying structures with intramolecular hydrogen bond, as hydrogen atoms in protein−ligand structures are typically not resolved. Searches were performed only for a subset of interactions, namely, six- and seven-membered hydrogen bond motifs of C=O acceptor with NH and OH donors. Slightly modified search queries had to be used: distance donor···acceptor $d \leq 3.0$ Å, angle base-donor···acceptor $70° \leq \alpha_1 \leq 180°$, and angle donor···acceptor-base $70° \leq \alpha_2 \leq 180°$. NH and OH donors were defined as nitrogen and oxygen atoms that are connected by single bonds to two and one neighbor atoms, respectively. Apart from these different geometric and donor type definitions, queries were run in identical manner in the CSD and the PDB ligand database. The propensity of a particular substructure to form an intramolecular hydrogen bond is expressed as the ratio between the number of entries that fulfill these geometric criteria and the total number of entries in the database containing this substructure. Certain topologies cannot possibly adopt a conformation required for intramolecular hydrogen bonding. These were eliminated by requiring that the torsion angles about any cyclic or acyclic double bond be within the interval −90° to 90°. Furthermore, we removed trivial solutions in which the number of consecutive cyclic bonds between donor and acceptor is ≥(total number of linker bonds − 1), as these have only 1 degree of freedom at best.

We report intramolecular hydrogen bond motifs that show more than five hits in the CSD database and contain a substructure that is represented more than five times in Prous. We distinguish between Csp2 (C3), Csp3 (C4), N, and O atom types as well as cyclic (c) and acyclic (a) bond types. The linker between donor and acceptor is abbreviated by the sequence of bond and atom types; i.e., cC3aC3a stands for a linker that connects hydrogen bond acceptor and donor by cyclic-Csp^2-acyclic-Csp^2-acyclic bond and atom types.

Conformational Energy Calculations. For model compounds **1d**, **2d**, **3b**, and **4b**, a systematic conformational search was performed with MOE[19] (version 2008.10) using the MMFF94x force field and without nonbonded interaction cutoff. Also cis-amides were included in the sampling and the resulting conformations minimized with default settings prior to analysis. Conformational searches were performed both in vacuo and in water (generalized Born solvation model). The conformations with lowest molecular mechanical energy of open and closed form each were selected for subsequent quantum mechanical optimizations with Jaguar[20] using the density functional B3LYP and a 6-31G** basis set. Both conformational energy differences were calculated in the gas phase and in the Minnesota water solvation model SM8.

Synthesis of Model Compounds. Amides **1a**−**d** were prepared from the commercially available precursor acids by transforming them into the corresponding acid chlorides with oxalyl chloride/cat. DMF in toluene at ambient temperature followed by reaction with either methylamine in THF (2 M) or dimethylamine in THF (2 M). 1-Methyl-3-phenylurea **2c** was synthesized by reacting aniline with methyl isocyanate in DMF at room temperature overnight, whereas 1,1-dimethyl-3-phenylurea

2a was obtained by stirring aniline with dimethylcarbamoyl chloride in the presence of triethylamine in CH_2Cl_2.[21] 1,1-Dimethyl-3-pyridin-2-ylurea **2b** was synthesized pursuing the same method (alternatively with Hünig's base instead of triethylamine); however, because of the ambidentic nature and the low reactivity the yield was extremely low (8%), and scrupulous purification was necessary to get the product in pure form. 1,3-Dipyridin-2-ylurea was formed as typical side product. 1-Methyl-3-pyridin-2-ylurea **2d**, finally, was prepared in decent yield by reacting 2-aminopyridine with methyl isocyanate again in DMF at ambient temperature. *N*-(1-Methyl-1*H*-benzoimidazol-2-yl)acetamide **3a** was produced by acetylation of commercially available 1-methyl-1*H*-benzoimidazol-2-ylamine with acetyl chloride/pyridine in CH_2Cl_2.[22] Extensive NMR spectroscopy was necessary to prove that the acetyl group is attached to the endocyclic nitrogen; reaction with acetic anhydride (neat) led to the identical product. We note in passing that **3a** exists as tautomeric mixture; as a consequence thereof, it was not possible to record a well resolved ^{13}C NMR. **3b**, on the other hand, was similarly synthesized from 1*H*-benzoimidazol-2-ylamine with neat acetic anhydride.[23] *N*-(1*H*-Benzoimidazol-2-ylmethyl)acetamide **4b** was bought from a commercial supplier (Princeton), whereas **4a**, finally, was obtained from *C*-(1*H*-Benzoimidazol-2-yl)methylamine by acetylation with acetyl chloride/pyridine in CH_2Cl_2. All compounds were purified by chromatography and/or crystallization to a purity of at least 95% as determined by HPLC. Full spectroscopic characterization of all model compounds is provided in the Supporting Information (pp S2−S48).

Molecular Property Measurements and Calculations. All compounds were tested with respect to their physicochemical behavior. The log *D* values were determined by the new CAMDIS method.[24] Kinetic solubilities were measured according to the LYSA protocol and are reported in μg/mL.[25] Permeation constants were obtained from the PAMPA assay[26] and are expressed in 10^{-6} cm/s. Details of each experimental setup are given in the Supporting Information (pp S49−S50). Lipophilicity was calculated with the program ClogP, version 4.94.[27]

Results and Discussion

In the following, we first discuss the statistical trends observed for intramolecular hydrogen bonds forming five- to eight-membered rings as observed in crystallographic databases. In the second part of the paper, we analyze molecular property differences of specifically designed sets of model compounds in which one molecule can form an intramolecular hydrogen bond while one or more closely related structures cannot.

1. Analysis of Crystallographic Data. The probability of intramolecular hydrogen bond formation is far greater for

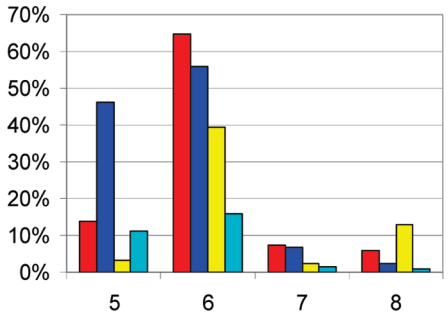


Figure 3. CSD-derived probabilities for the formation of intramolecular hydrogen bonds of ring sizes 5−8 with NH as hydrogen bond donor and different hydrogen bond acceptors: C=O (red), heterocyclic N acceptor (blue), S=O (yellow), alkoxy (cyan).

Table 1. CSD Hit Statistics of Intramolecular Hydrogen Bond Formation of Ring Sizes 5−7 for Different Combinations of Hydrogen Bond Acceptors (Carbonyl, Disubstituted N in Heterocycle) and Donors (NH, OH)[a]

	NH	OH
	Five-Membered	
C=O	13.7	16.8
N	46.1	11.5
	Six-Membered	
C=O	64.7	48.6
N	56.0	42.4
	Seven-Membered	
C=O	7.4	9.0
N	6.7	10.8

[a] Numbers indicate the percentage of entries with intramolecular hydrogen bond (% hb).

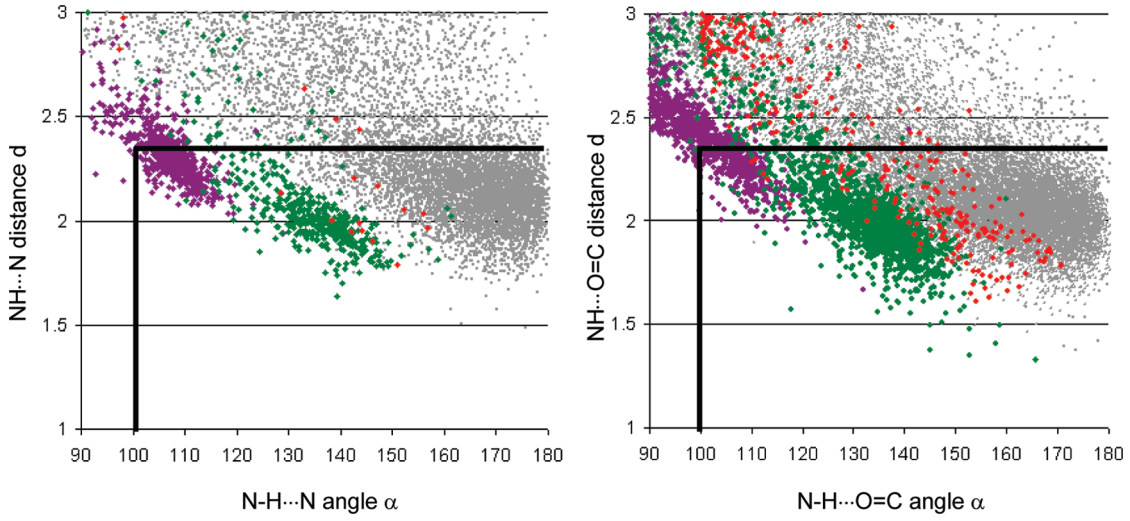


Figure 4. Plot of hydrogen bond distance vs angle for NH···N and NH···O=C. Ring sizes 5 (magenta), 6 (green), and 7 (red) are plotted along with the distribution for unconstrained intermolecular hydrogen bonds (gray). The geometry constraints used in this study are plotted as thick lines.

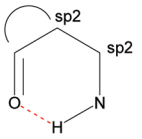
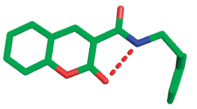
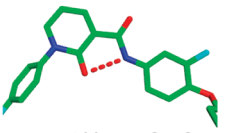
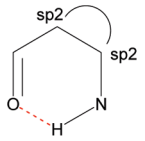
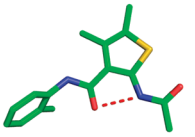
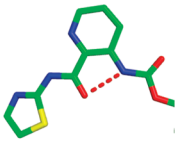
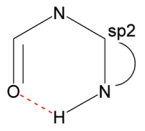
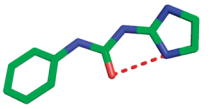
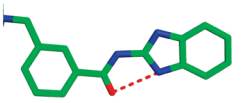
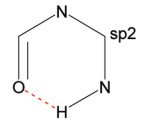
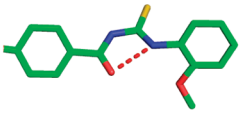
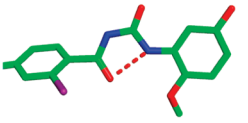
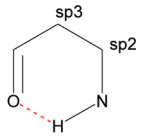
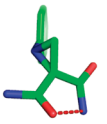
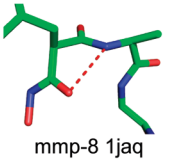
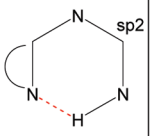
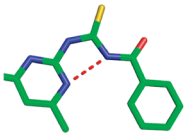
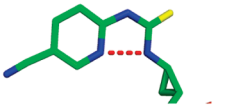
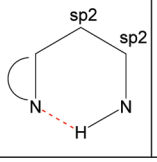
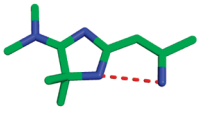
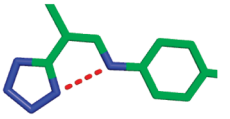
Topology		% hb	CSD example	PDB example
cC3aC3a		93.0	 ACEMEB	 c-met kinase 3ce3
aC3cC3a		89.5	 ABEKIC	 met-aminopeptidase 2nq6
aNaC3c		93.5	 CIVRUV	 hiv protease 1bwb
aNaC3a		85.3	 ABEFAP	 glyco-phosphorylase 2ati
aC4aC3a		17.5	 COYMOS	 mmp-8 1jaq
aNaC3a		80.9	 AWUBID	 hiv rev-transcriptase 1iky
aC3aC3a		66.7	 BOJYII	 beta-lactamase 1pzp

Figure 5. Selected topologies, CSD hit statistics, and crystal structure examples of intramolecular NH···O=C and NH···N interactions in six-membered rings. The topology abbreviation lists the sequence of bond and atom types in the linker connecting hydrogen bond acceptor with donor. C3 and C4 are sp^2 and sp^3 carbon atoms, and a and c denote acyclic and cyclic bond types, respectively. Statistics for other six-membered ring topologies are supplied in the Supporting Information (pp S51−S55).

six-membered rings than for any other system (Figure 3). Here, the probability of intramolecular ring formation correlates with the acceptor strength in the order C=O > heterocyclic N acceptor > S=O > alkoxy.[28] In all other cases, specific substructures with a particularly high propensity of hydrogen bond formation dominate the statistics. These will be discussed in the individual paragraphs below.

Geometric parameters of all small-ring intramolecular hydrogen bonds lie outside the typically observed ranges for unconstrained systems. Figure 4 illustrates this for carbonyl and nitrogen acceptors. A linear correlation between bond length and angle is observed in all cases. Intermolecular hydrogen bonds are close to linear, preferentially with angles greater than 150°. Only seven-membered ring intramolecular hydrogen bonds come close to these values. Six-membered rings have angles between 130° and 140° but the same distances as intermolecular hydrogen bonds. Five-membered rings have longer distances and smaller angles just within the cutoffs applied here (decreasing the lower cutoff to 75° changes the statistics only marginally by ~1%). On these grounds it may be argued that five-membered rings should not be regarded as classical hydrogen bonds but, more general, as favorable electrostatic interactions. We will still cover them here, since the

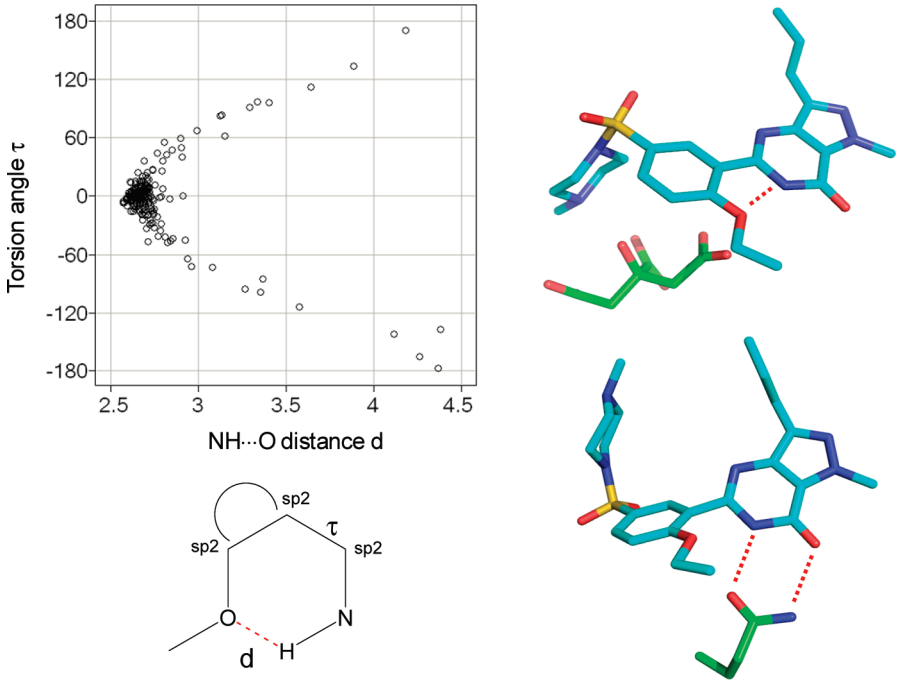


Figure 6. Left: Plot of torsion angle vs hydrogen bond distance for arylether oxygen atoms interacting with an NH group separated by three linker atoms. Intramolecular hydrogen bonds are not uncommon in drugs. Top right: Small-molecule crystal structure of the citrate salt of sildenafil revealing an intramolecular hydrogen bond (CSD entry FEDTEO). Bottom right: Sildenafil in complex with phosphodiesterase 5a with intermolecular hydrogen bonds only (PDB code 1tbf).

conformation−property relationships for larger rings described in this article are applicable to such systems as well.

Table 1 allows a comparison between the two donor types (NH and OH) and the two most important acceptor types (C=O and cyclic N). As observed by Bilton et al.,[14] the overall propensities are quite similar for the two donors. An exception is unusually frequent NH···N five-membered ring systems, dominated by amide-substituted heterocycles adopting a high propensity planar conformation. There is also no general difference between nitrogen and carbonyl as an acceptor, indicating that formal lone pair directionality is not correlated with the likelihood of hydrogen bond formation.

Six-Membered Rings. Individual topologies, hit statistics, and illustrative examples of selected interactions are summarized in Figure 5. Detailed statistics for all other systems is supplied in the Supporting Information (pp S51−S55). All substructures with a high probability of hydrogen bond formation are variations on a common theme: They all share two sp^2-hybridized linker atoms (amides, heteroaromatic rings) leading to planar, conjugated systems. The high stability of this type of intramolecular hydrogen bond has been termed resonance-assisted hydrogen bonding and can be rationalized by enhanced π-delocalization.[29] Substructures containing sp^3 atoms in the linker do not profit from resonance assistance and are conformationally more flexible resulting in a preference for the open form (e.g., aC4aC3a). Further constraints in the form of cyclic linker bonds slightly increase the propensity for internal hydrogen bond formation, as illustrated by the comparison of the topologies aNaC3c and aNaC3a. The general observation that carbonyl and nitrogen acceptors have a similar propensity for hydrogen bond formation also holds for specific linkers. For example, the aNaC3a linker with the N acceptor occurs in the hydrogen-bonded form in 81% of the entries and with C=O as acceptor in 85% of the entries.

Queries on the PDB confirm the preference for intramolecular hydrogen bond formation in the case of sp^2-hybridized linkers, albeit with lower propensity (Supporting Information p S56). For nonplanar rings the quantitative agreement between CSD and PDB results is higher, indicating that the increased probability for planar hydrogen bonded systems in small molecule crystals is caused by packing effects.

Slightly less pronounced preferences for resonance-assisted hydrogen bonds are found with weaker acceptors such as alkoxy groups bound to aromatic rings (Supporting Information p S55). CSD statistics show that the planar hydrogen bonded conformations prevail in the solid state, but a significant number of open forms exist as well (Figure 6). The vasodilator sildenafil[30] is an example of such a system. When complexed with its target, phosphodiesterase 5a, the lactame unit, is engaged in an intermolecular hydrogen bond with a Gln side chain.[31] When crystallized as its citrate salt, sildenafil adopts a planar conformation with the pyrimidinone NH interacting intramolecularly with the ethoxy group (Figure 6).[32]

The general observation that the propensity for intramolecular hydrogen bonding correlates with increasing hydrogen bond acceptor strength is nicely illustrated by the preferred conformations of ester groups interacting with NH donors (Figure 7). Because of its higher acceptor strength, the C=O group of esters is found considerably more often in the syn than in the anti arrangement (CSD ratio of ~20:1).

The high propensity to form resonance-assisted hydrogen bonds has been used in the design of new kinase inhibitors.[6] On the basis of a known bicyclic kinase inhibitor scaffold, pyrimidin-4-ylureas were suggested as bioisosteres

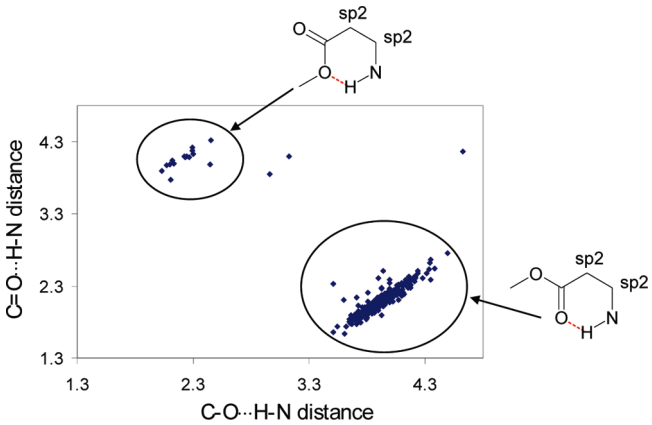


Figure 7. CSD-derived relative preferences of alkoxy and carbonyl functionalities of ester groups to engage in six-membered resonance-assisted intramolecular hydrogen bonds with NH donors.

(Figure 8a). Molecules containing this substructure indeed turned out to be inhibitors of multiple kinases, and the binding mode was confirmed by a cocrystal structure. The strong preference for intramolecular hydrogen bond formation of urea-substituted pyri(mi)dines has also been confirmed by NMR spectroscopy[33] and by model calculations on adenosine receptor agonists.[34] Hydrogen bonds of this type were shown to persist in water.[35] In our studies, we find a probability of 81% that intramolecular hydrogen bond formation occurs for such a topology (Figure 5, N···HN, aNaC3a). The literature contains a number of further examples where six-membered hydrogen bonded pseudorings effectively mimic an aromatic ring, offering a similar extent of π-delocalization and pharmacophore features (Figure 8b−d), but overall this possibility of "scaffold hopping" still seems underexplored.[4,5,7,8,36]

Seven-Membered Rings. In contrast to six-membered rings, planar, resonance-assisted hydrogen bonds are considerably less likely for seven-membered rings because of steric repulsion. In fact, we could identify only one substructure that fulfills these criteria (and it is rare in druglike compounds). As shown in Figure 9, two amide substituents in the ortho position off a five-membered ring heterocycle are in plane with each other with a propensity of 60% (11 hits in 18 CSD structures). No analogous examples with intramolecular hydrogen bonds are known for six-membered aromatic rings in the linker (23 CSD entries), most likely because of higher angle strain.

In the seven-membered ring C=O···HN structures (Figure 10), donor and acceptor are not in plane; the acceptor O atom can be as far as 1.8 Å above the donor plane. Specific topologies occur quite frequently in the ring-closed form. Strikingly, the aC4aC3cC3a ring system (first entry in Figure 10) is homologous to cyclohepta-1,4-diene in terms of electronic structure and geometry. One may regard the carbonyl group and the aromatic bond participating in the seven-membered ring as double bond equivalents. The conformation of this ring as exemplified by MISBEW is a "folded" structure, consisting of two sets of five and four coplanar atoms, corresponding to the lowest energy conformer of cyclohepta-1,4-diene.[37]

The statistics for N-Csp3-Csp2 linkers connecting C=O acceptor and NH donor is dominated by peptide structures. The cyclic conformation, which corresponds to the γ-turn,[38] is rare: In the CSD, only 5 out of 1138 CSD entries (0.4%)

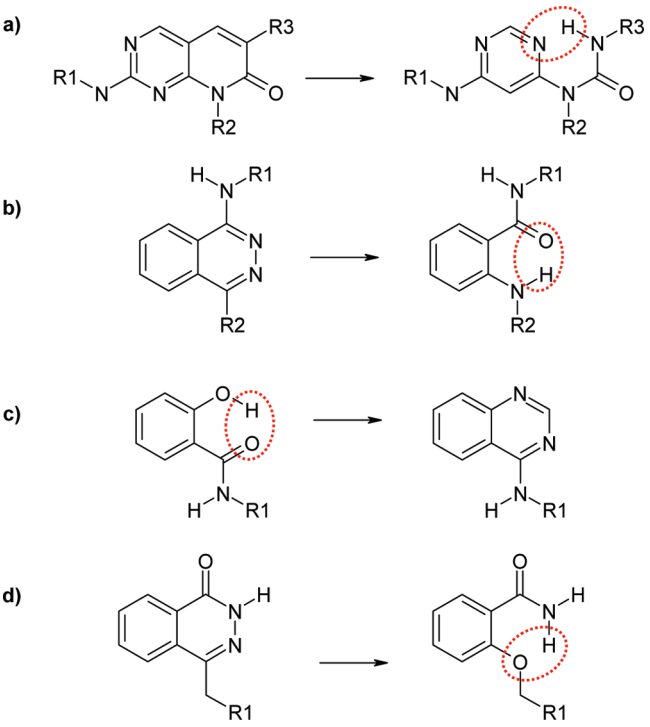


Figure 8. Examples of designed scaffold replacements comprising an intramolecular hydrogen bond: (a) pyrimidin-4-ylurea kinase inhibitors;[6] (b) anthranilamide kinase inhibitors;[5] (c) 4-aminoquinazoline scytalone dehydratase inhibitors;[4] (d) alkoxybenzamide poly(ADP-ribose) polymerase inhibitors.[7] Intramolecular hydrogen bonds are highlighted in red.

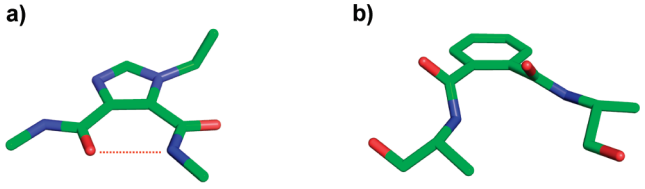


Figure 9. (a) Planar intramolecular seven-membered ring system involving five-membered heterocycles (BEZYAG). (b) No intramolecular hydrogen bonds are observed for systems involving six-membered rings (LICKIS).

form an intramolecular hydrogen bond. In the PDB, 75 of 1786 ligands (4%) form γ-turns. It might be argued that the more lipophilic protein environment promotes the formation of the less polar hydrogen bonded form ("hydrophilic collapse"). Many PDB structures are not pure γ-turns but intermediate structures between β sheet and cyclic structure, with bifurcated inter- and intramolecular hydrogen bonds.

We are not aware of examples where seven-membered ring intramolecular hydrogen bonds have been designed in medicinal chemistry. Where they were observed, they had not been predicted by modeling.[39]

Eight-Membered Rings. The overall frequency of eight-membered ring intramolecular hydrogen bonds declines further; the number of alternative conformations is even larger here. Two substructures still stand out: intramolecular S=O···NH hydrogen bonds are more frequent in eight- than in seven-membered rings. This is in particular true for Csp2-Csp2-Csp3-Csp3 linkers, even if there are few examples (6 CSD hits in 10 entries, Figure 11a). Analogous systems

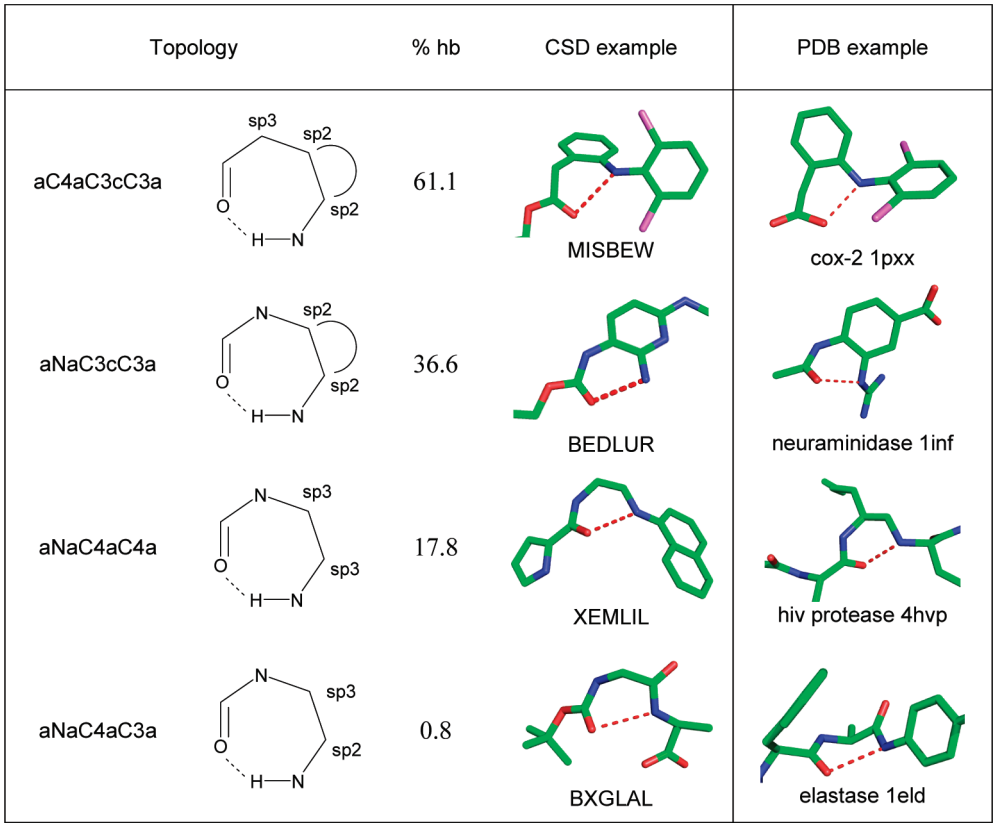


Figure 10. Topologies, CSD hit statistics, and crystal structure examples for the C=O···NH interaction in seven-membered rings. Statistics for other seven-membered ring topologies are supplied in the Supporting Information (pp S51−S53).

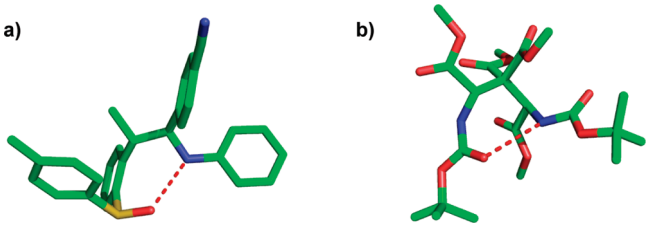


Figure 11. Examples for eight-membered rings with relatively high propensity: (a) S=O···NH interaction with topology aC3cC3a-C4aC4a (DIGNAJ); (b) C=O···NH interaction with topology aNaC4aC4aC4a (WUGJIR).

with C=O acceptors have a lower propensity for this arrangement and are characterized by longer hydrogen bond distances. Apparently, the special geometry of the S=O group is well suited to engage in this type of intramolecular hydrogen bond. The second substructure is C=O···NH contacts connected by a N-Csp3-Csp3-Csp3 linker (6 CSD hits in 21 entries, Figure 11b).

Five-Membered Rings. For substructures where a donor is separated by three bonds from an acceptor, we often find a strong preference for planar geometries that is not observed when the acceptor is replaced by carbon. This difference in conformation is caused by a combination of reduced steric repulsion and favorable electrostatic interaction between H-bond donor and acceptor. This effect is illustrated with torsion histograms for two frequently occurring motifs, pyridine-2-amides and alkoxyacetamides, and their corresponding carbon analogues (Figure 12). These examples are representative of the many planar NH···N

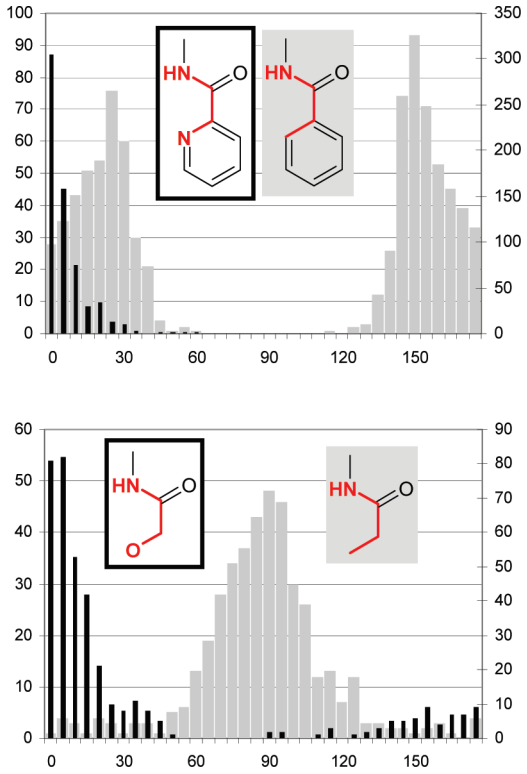


Figure 12. Torsion histograms for two analogue systems in which the planar motif (black) contains a hydrogen bond acceptor that can interact with the NH donor, while the corresponding carbon congener is bent (gray).

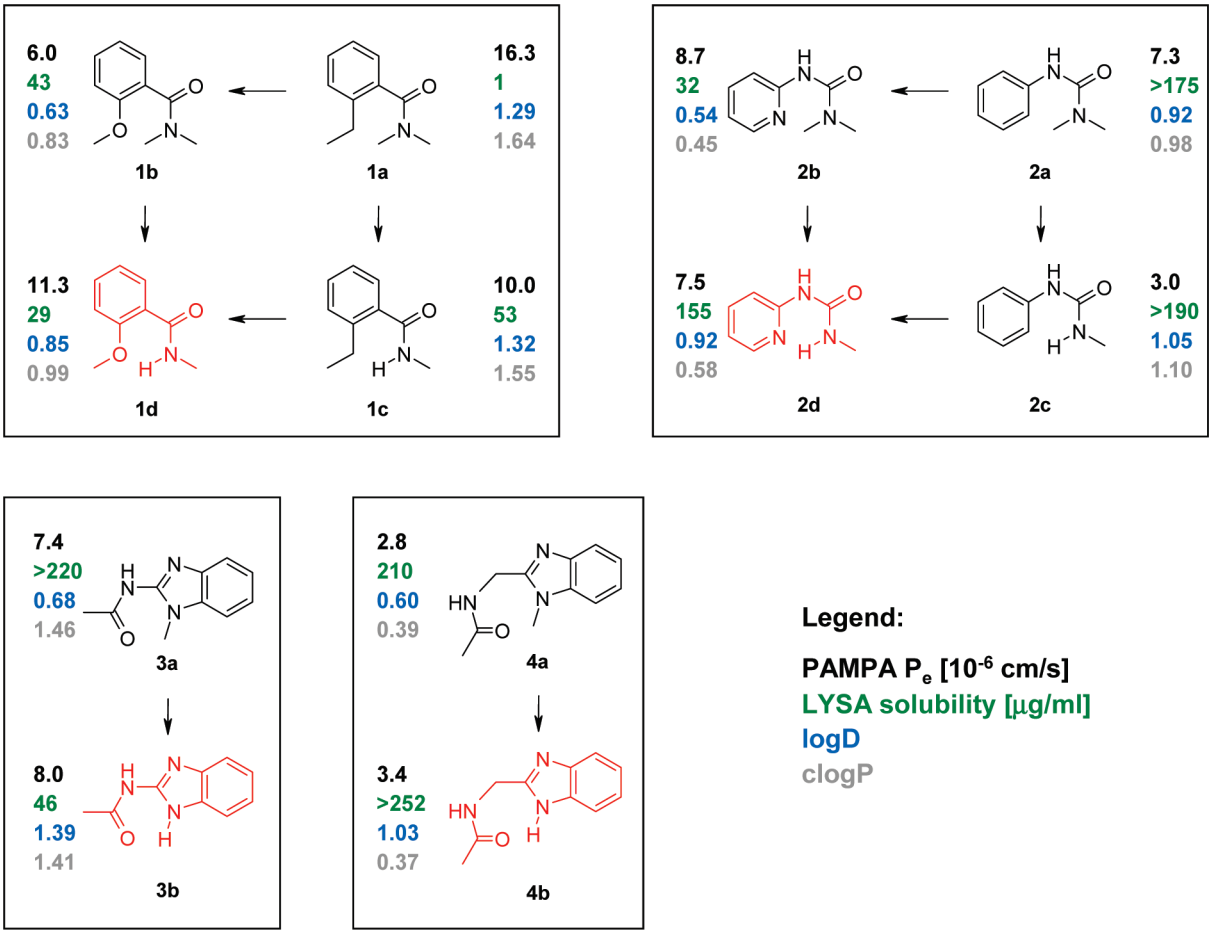


Figure 13. Experimental and calculated molecular properties of four different model systems. Molecules **1d** (2-methoxy-*N*-methylbenzamide), **2d** (1-methyl-3-pyridin-2-ylurea), **3b** (*N*-(1*H*-benzoimidazol-2-yl)acetamide), and **4b** (*N*-(1*H*-benzoimidazol-2-ylmethyl)acetamide) are able to form intramolecular hydrogen bonds and are highlighted in red. Experimental properties are PAMPA permeation constant P_e (black), LYSA water solubility (green), and octanol/water partition coefficient log *D* (blue). clogP (gray) is calculated log *P*.

and NH···alkoxy arrangements observed in the CSD (see Supporting Information p S57 for crystal structure examples). Such motifs occur very frequently in development compounds.

2. Analysis of Molecular Properties. The formation of an intramolecular hydrogen bond should effectively remove one donor and one acceptor function from the surface of a molecule and thus result in increased lipophilicity and membrane permeability and in reduced aqueous solubility. To study this hypothesis in detail, we synthesized and characterized four sets of closely related chemical structures. In each set, only one structure is able to form an intramolecular hydrogen bond; the others lack the necessary hydrogen bond donor or acceptor. Figure 13 gives an overview of the structures and their respective properties. We measured the passive permeability (PAMPA) permeation constant P_e, the octanol/water partition coefficient log *D*, and kinetic solubility. In addition, the calculated lipophilicity clogP is contrasted with the experimental log *D* values.

Molecular property changes of the benzamides **1a**−**d** follow the expected trends. Starting from **1a**, introduction of a weak acceptor (**1b**) or a donor (**1c**) decreases permeability and increases solubility. The presence of both functional groups in **1d** has the opposite effect. The log *D* value also decreases from **1a** to **1b**. The log *D* values of **1a** and **1c** are identical in spite of the introduction of the donor. This is not readily explained on the basis of the two-dimensional structures alone. Most likely the pronounced out-of-plane rotation of the tertiary amide substituent affects distribution behavior.

The benzimidazoles **3a** and **3b** also follow the above trends. Removal of a methyl group in **3a** leads to **3b**, which has a higher log *D* value, slightly higher permeability, and decreased solubility. Although we do observe the expected trends here, it is noted that **3a** is in equilibrium with a second tautomer, which results from a hydrogen shift of the amide NH to the nonmethylated benzimidazole nitrogen. This tautomer can form a resonance-assisted six-membered intramolecular hydrogen bond (C=O···HN) of topology aNaC3c. The balance between the two tautomers very likely depends on the nature of the solvent. We have not performed further analyses here.

It is instructive to compare clogP values and experimental log *D* values for the four model systems (Figure 13). Not unexpectedly, clogP calculations underestimate the lipophilicity of molecules with intramolecular hydrogen bonds because the underlying fragment-based approach does not take into account internal pairing of polar groups. This is most pronounced in the pair **3a** and **3b**. The two molecules are predicted to have equal lipophilicity (clogP of 1.46 vs 1.41), whereas the measured log *D* value of the hydrogen bonded analogue **3b** is 0.71 units higher than that of **3a**. On average, it seems that clogP predictions should be increased

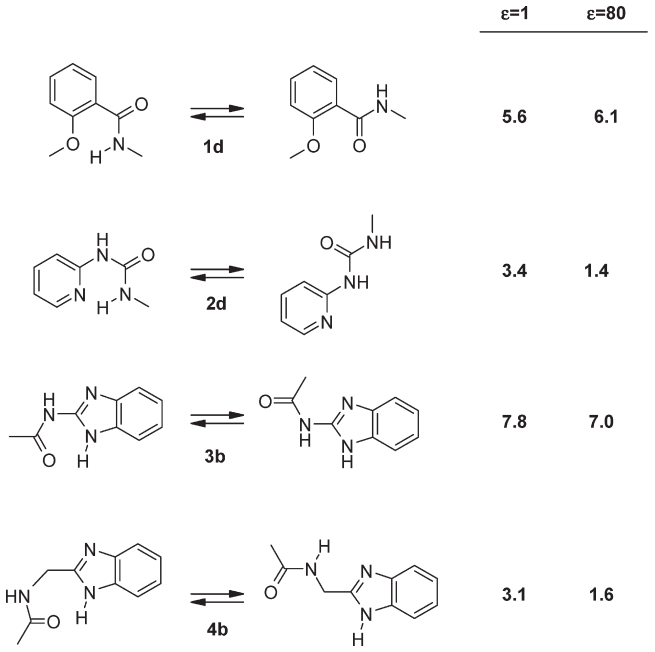


Figure 14. Quantum mechanically (B3LYP/6-31G**) calculated energy differences between open and closed conformations. Differences are given in kcal/mol for gas phase (dielectric constant $\varepsilon = 1$) and water environment ($\varepsilon = 80$). Hydrogen bonded 3D conformations are shown: **1d** (analogue of Figure 6, top right), **2d** (Figure 5, aNaC3a), **3b** (Figure 5, aNaC3c), **4b** (analogue of Figure 10, aNaC4aC3a).

by a correction factor of 0.4 for each intramolecular hydrogen bond. Note that polar surface area calculations of molecules with intramolecular hydrogen bonds are overestimates for the same reason.

In contrast to the structures **1** and **3**, the molecular property profiles of the urea-substituted aryls **2** and the benzimidazoles **4** deviate from the above simple trends. In both cases, the structures with a potential intramolecular hydrogen bond exhibit unexpectedly high aqueous solubility and, for the urea structures, a less consistent trend in log *D* values.

The unexpectedly high solubility of **2d** and **4b** might be due to a different equilibrium between the open (polar) and hydrogen-bonded (unpolar) conformations in aqueous solution than in the cases of **1d** and **3b**. To investigate this hypothesis, we performed high-level quantum mechanical calculations of conformational energy differences of the lowest open and closed forms of the respective hydrogen bonded representatives **1d**, **2d**, **3b**, and **4b** in both a low- and high-dielectric environment (Figure 14). In the low dielectric environment, the energy difference between the open and closed forms is prohibitively high in all cases. In the high dielectric, however, the energy difference stays high for **1d** and **3b** but decreases to about 1.5 kcal/mol for **2d** and **4b**. Taken at face value, this energy difference corresponds to a population of approximately 10% of the higher energy conformer in aqueous solution. We interpret the energy differences as a trend only, indicative of a significant population of the open conformation in aqueous solution positively affecting solubility.

The lesson to be learned from systems **2** and **4** is that the introduction of a structural feature facilitating intramolecular hydrogen bond formation can be beneficial for the overall property profile, leading to increased solubility, enhanced permeability, and only a modest increase in lipophilicity. The prerequisite for such an effect seems to be that the polar, open conformation should be a low-energy conformation in water. The benzamide **1d** also shows attractive overall properties and suffers only from slightly reduced solubility. In this case, crystallographic data (Figure 6) indicate that the intramolecular hydrogen bond is rather weak, leaving the amide torsion rather flexible. This leaves us with two alternative strategies for the introduction of intramolecular hydrogen bonds with a beneficial overall profile: choosing a strong hydrogen bond with a distinct alternative open conformation accessible in water or choosing a relatively flexible and weak hydrogen bond with a continuum of open and closed states.

Conclusions

By means of database searches in the CSD, we have identified a considerable number of substructures with the propensity to form intramolecular hydrogen bonds. We found several indicators that intramolecular hydrogen bonding is correlated with acceptor strength. The highest frequency of intramolecular hydrogen bonds have planar, six-membered rings stabilized by conjugation with a π-system. Such substructures have been actively applied in drug design. A variety of further, far less explored topologies have been identified: Weaker six-membered ring hydrogen bonds containing one sp^3 center and, in particular, a number of nonplanar seven-membered and eight-membered ring topologies. Five-membered ring intramolecular hydrogen bonds have the smallest angles and the longest distances and are just within the geometric limits of a hydrogen bond as defined here. Nevertheless, because of their planarity, these topologies can be useful modules in conformation design.

The molecular properties measured for four specifically designed model systems did not reveal a uniform behavior with respect to membrane permeability, solubility, and lipophilicity. To understand the trends within these systems, it is important to consider the equilibrium between the polar open conformations and the unpolar closed conformation both in aqueous solution. Simultaneous high solubility and permeability can be achieved when the open conformation is of sufficiently low energy to be significantly populated in water. This can be predicted to some extent by quantum mechanical calculations of conformational energies in a high dielectric environment. The targeted design of intramolecular hydrogen bonds also requires an assessment of hydrogen bond interaction strength. Weaker intramolecular hydrogen bonds avoid an extreme shift of the equilibrium to the closed form. In that respect, potential design strategies are to reduce the hydrogen bond acceptor or donor strengths[28] or to capitalize on nonplanar seven- instead of planar six-membered rings. Finally, introduction of steric strain in the closed conformation may be a useful strategy to affect the equilibrium. In seven-membered ring topologies similar to Figure 9a (pyrazoles instead of imidazoles), we have found drastically improved aqueous solubility when the amide nitrogen not involved in intramolecular hydrogen bonding is fully substituted. X-ray structures showed an intramolecular hydrogen bond for both primary and tertiary amides but very different torsional angles of the two amide planes of 30° and 70°, respectively (results not shown).

Intramolecular hydrogen bonds are abundant in medicinal chemistry; one will find examples in any recent issue of this

journal.[40,41] We are confident that our analysis of topologies and properties of this interaction type provides a useful basis for its more directed and rational use.

Acknowledgment. Synthesis of model compounds by Stefan Bürli, exhaustive spectroscopic characterization by Christian Bartelmus, Markus Bürkler, Dr. Joseph Schneider, and Dr. Monira Siam, and physicochemical measurements by Isabelle Parrilla, Frank Senner, and Björn Wagner are gratefully acknowledged.

Supporting Information Available: Synthesis and characterization data for model compounds (pp S2−S48); methods for measurement of physicochemical properties (pp S49−S50); detailed CSD statistics for C=O···H−N, C=O···H−O, N···H−N, N···H−O, and alkoxy···H−N interactions (pp S51−S55); a comparison between CSD and PDB statistics of C=O···H−N and C=O···H−O interactions (p S56); crystal structure examples of N···H−N and alkoxy···H−N interactions in five-membered rings with particularly high propensities (p S57). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Nomura, M.; Kinoshita, S.; Satoh, H.; Maeda, T.; Murakami, K.; Tsunoda, M.; Miyachi, H.; Awano, K. (3-Substituted benzyl)thiazolidine-2,4-diones as structurally new antihyperglycemic agents. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 533–538.
(2) Harter, W. G.; Albrect, H.; Brady, K.; Caprathe, B.; Dunbar, J.; Gilmore, J.; Hays, S.; Kostlan, C. R.; Lunneya, B.; Walker, N. The design and synthesis of sulfonamides as caspase-1 inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 809–812.
(3) Van Zandt, M. C.; Sibley, E. O.; McCann, E. E.; Combs, K. J.; Flam, B.; Sawicki, D. R.; Sabetta, A.; Carrington, A.; Sredy, J.; Howard, E.; Mitschler, A.; Podjarny, A. D. Design and synthesis of highly potent and selective (2-arylcarbamoyl-phenoxy)-acetic acid inhibitors of aldose reductase for treatment of chronic diabetic complications. *Bioorg. Med. Chem.* **2004**, *12*, 5661–5675.
(4) Hodge, C. N.; Pierce, J. A diazine heterocycle replaces a six-membered hydrogen-bonded array in the active site of scytalone dehydratase. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1605–1608.
(5) Furet, P.; Bold, G.; Hofmann, F.; Manley, P.; Meyer, T.; Altmann, K.-H. Identification of a new chemical class of potent angiogenesis inhibitors based on conformational considerations and database searching. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2967–2971.
(6) Furet, P.; Caravatti, G.; Guagnano, V.; Lang, M.; Meyer, T.; Schoepfer, J. Entry into a new class of protein kinase inhibitors by pseudo ring design. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 897–900.
(7) Menear, K. A.; Adcock, C.; Alonso, F. C.; Blackburn, K.; Copsey, L.; Drzewiecki, J.; Fundo, A.; Le Gall, A.; Gomez, S.; Javaid, H.; Lence, C. F.; Martin, N. M. B.; Mydlowski, C.; Smith, G. C. M. Novel alkoxybenzamide inhibitors of poly(ADP-ribose) polymerase. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3942–3945.
(8) Lord, A.-M.; Mahon, M. F.; Lloyd, M. D.; Threadgill, M. D. Design, synthesis, and evaluation in vitro of quinoline-8-carboxamides, a new class of poly(adenosine-diphosphate-ribose)polymerase-1 (PARP-1) inhibitor. *J. Med. Chem.* **2009**, *52*, 868–877.
(9) Ashwood, V. A.; Field, M. J.; Horwell, D. C.; Julien-Larose, C.; Lewthwaite, R. A.; McCleary, S.; Pritchard, M. C.; Raphy, J.; Singh, L. Utilization of an intramolecular hydrogen bond to increase the CNS penetration of an NK1 receptor antagonist. *J. Med. Chem.* **2001**, *44*, 2276–2285.
(10) Sasaki, S.; Cho, N.; Nara, Y.; Harada, M.; Endo, S.; Suzuki, N.; Furuya, S.; Fujino, M. Discovery of a thieno[2,3-*d*]pyrimidine-2, 4-dione bearing a *p*-methoxyureidophenyl moiety at the 6-position: a highly potent and orally bioavailable non-peptide antagonist for the human luteinizing hormone-releasing hormone receptor. *J. Med. Chem.* **2003**, *46*, 113–124.
(11) Rezai, T.; Bock, J. E.; Zhou, M. V.; Kalyanaraman, C.; Lokey, R. S.; Jacobson, M. P. Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.* **2006**, *128*, 14073–14080.
(12) Infantes, L.; Motherwell, W. D. S. Hydrogen bond competition between chemical groups: new methodology and the Cambridge Structural Database. *Z. Kristallogr.* **2005**, *220*, 333–339.
(13) Allen, F. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B* **2002**, *58*, 380–388.
(14) Bilton, C.; Allen, F. H.; Shields, G. P.; Howard, J. A. K. Intramolecular hydrogen bonds: common motifs, probabilities of formation and implications for supramolecular organization. *Acta Crystallogr. B* **2000**, *56*, 849–856.
(15) http://www.rcsb.org/pdb/.
(16) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
(17) http://www.desertsci.com/.
(18) Prous Science, S. A. Prous Science Integrity. http://integrity.prous.com
(19) http://www.chemcomp.com/.
(20) http://www.schrodinger.com/.
(21) Woods, W. G.; Crawford, R. F. 1,1-Dialkyl-3-(pyridyl)uea Hercides. U.S. Patent 3330641, **1967**.
(22) Pilyugin, V. S.; Sapozhnikov, Y. E.; Davydov, A. M.; Chikisheva, G. E.; Vorob'eva, T. P.; Klimakova, E. V.; Kiseleva, G. V.; Kuznetsova, S. L.; Davletov, R. D.; Sapozhnikova, N. A.; Yumadilov, R. K. ^{13}C NMR spectra and biological activity of *N*-(1*H*-benzimidazol-2-yl)benzamides. *Russ. J. Gen. Chem.* **2006**, *76*, 1653–1659.
(23) Terzioglu, N.; van Rijn, R. M.; Bakker, R. A.; De Esch, I. J. P.; Leurs, R. Synthesis and structure−activity relationships of indole and benzimidazole piperazines as histamine H4 receptor antagonists. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5251–5256.
(24) Fischer, H.; Kansy, M.; Wagner, B. Determination of High Lipophilicity Values. U.S. Patent 2006211121, European Patent 1705474, 2006.
(25) *Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability*; Van de Waterbeemd, H., Lennernas, H., Artursson, P., Eds.; Methods and Principles in Medicinal Chemistry, Vol. 18; Wiley-VCH: Weinheim, Germany, 2003; 579 pp.
(26) Kansy, M.; Avdeef, A.; Fischer, H. Advances in screening for membrane permeability: high-resolution PAMPA for medicinal chemists. *Drug Discovery Today: Technol.* **2004**, *1*, 349–355.
(27) BioByte Corp. http://www.biobyte.com.
(28) Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J.-Y.; Renault, E. The pKBHX database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J. Med. Chem.* **2009**, *52*, 4073–4086.
(29) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. Evidence for resonance-assisted hydrogen bonding from crystal-structure correlations on the enol form of the b-diketone fragment. *J. Am. Chem. Soc.* **1989**, *111*, 1023–8.
(30) Terrett, N. K.; Bell, A. S.; Brown, D.; Ellis, P. Sildenafil (Viagra), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1819–1824.
(31) Zhang, K. Y. J.; Card, G. L.; Suzuki, Y.; Artis, D. R.; Fong, D.; Gillette, S.; Hsieh, D.; Neiman, J.; West, B. L.; Zhang, C.; Milburn, M. V.; Kim, S.-H.; Schlessinger, J.; Bollag, G. A glutamine switch mechanism for nucleotide selectivity by phosphodiesterases. *Mol. Cell* **2004**, *15*, 279–286.
(32) Yathirajan, H. S.; Nagaraj, B.; Nagaraja, P.; Bolte, M. Sildenafil citrate monohydrate. *Acta Crystallogr. E* **2005**, *61*, 489–491.
(33) Sudha, L. V.; Sathyanarayana, D. N. Infrared and proton NMR study of molecular conformation of some *N*,*N*′-arylalkylureas. *J. Mol. Struct.* **1984**, *125*, 89–96.
(34) Ferretti, V.; Pretto, L.; Tabrizi, M. A.; Gilli, P. Role of strong intramolecular N−H···N hydrogen bonds in determining the conformation of adenosine-receptor ligands. *Acta Crystallogr. B* **2006**, *62*, 634–641.
(35) Jansma, A.; Zhang, Q.; Li, B.; Ding, Q.; Uno, T.; Bursulaya, B.; Liu, Y.; Furet, P.; Gray, N. S.; Geierstanger, B. H. Verification of a designed intramolecular hydrogen bond in a drug scaffold by nuclear magnetic resonance spectroscopy. *J. Med. Chem.* **2007**, *50*, 5875–5877.
(36) Zhu, G.-D.; Gandhi, V. B.; Gong, J.; Thomas, S.; Luo, Y.; Liu, X.; Shi, Y.; Klinghofer, V.; Johnson, E. F.; Frost, D.; Donawho, C.; Jarvis, K.; Bouska, J.; Marsh, K. C.; Rosenberg, S. H.; Giranda, V. L.; Penning, T. D. Synthesis and SAR of novel, potent and orally bioavailable benzimidazole inhibitors of poly(ADP-ribose) polymerase (PARP) with a quaternary methylene-amino substituent. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3955–3958.

(37) 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. B* **1994**, *50*, 395–404.
(38) Nemethy, G.; Printz, M. P. The gamma-turns, a possible folded conformation of the polypeptide chain. Comparison with the beta-turn. *Macromolecules* **1972**, *5*, 755–758.
(39) Foloppe, N.; Fisher, L. M.; Howes, R.; Kierstan, P.; Potter, A.; Robertson, A. G. S.; Surgenor, A. E. Structure-based design of novel Chk1 inhibitors: insights into hydrogen bonding and protein−ligand affinity. *J. Med. Chem.* **2005**, *48*, 4332–4345.
(40) Bikker, J. A.; Brooijmans, N.; Wissner, A.; Mansour, T. S. Kinase domain mutations in cancer: implications for small molecule drug design strategies. *J. Med. Chem.* **2009**, *52*, 1493–1509.
(41) Schroeder, G. M.; An, Y.; Cai, Z.-W.; Chen, X.-T.; Clark, C.; Cornelius, L. A. M.; Dai, J.; Gullo-Brown, J.; Gupta, A.; Henley, B.; Hunt, J. T.; Jeyaseelan, R.; Kamath, A.; Kim, K.; Lippy, J.; Lombardo, L. J.; Manne, V.; Oppenheimer, S.; Sack, J. S.; Schmidt, R. J.; Shen, G.; Stefanski, K.; Tokarski, J. S.; Trainor, G. L.; Wautlet, B. S.; Wei, D.; Williams, D. K.; Zhang, Y.; Zhang, Y.; Fargnoli, J.; Borzilleri, R. M. Discovery of *N*-(4-(2-amino-3-chloropyridin-4-yloxy)-3-fluorophenyl)-4-ethoxy-1-(4-fluorophenyl)-2-oxo-1,2-dihydropyridine-3-carboxamide (BMS-777607), a selective and orally efficacious inhibitor of the Met kinase superfamily. *J. Med. Chem.* **2009**, *52*, 1251–1254.