Data-mining the Diaryl(thio)urea Conformational Landscape: Understanding the Contrasting Behavior of Ureas and Thioureas with Quantum Chemistry

Guilian Luchini[‡], David M. H. Ascough[†], Juan V. Alegre-Requena[‡], Veronique Gouverneur[†], Robert S. Paton[*‡†]

[‡] Department of Chemistry, Colorado State University, Fort Collins, CO 80523, USA
[†] Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA, UK

robert.paton@colostate.edu

Abstract. The conformations adopted by urea and thiourea functional groups influence catalysis and binding. We combine data-mining with quantum chemical calculations to understand the differences in conformational behavior for these two important structural motifs. We developed a Python tool to automate the compilation of X-ray structural information and perform conformational clustering and visualization, based on SMILES input. While diarylureas have an overwhelming preference for the *anti,anti*-conformer, diarylthioureas adopt a mixture of *anti,anti*- and *anti,syn*-conformers. Computations show the *anti,anti*-thiourea conformer is destabilized by out-of-plane rotations which avoid a steric clash with the sulfur atom. These conformational preferences were studied computationally under a variety of conditions, and apart from in the gas-phase, a preference for *anti,anti*-ureas was found. Consistent with experiments, this preference increases in more polar environments. Quantitative predicted ratios are sensitive to the computational treatment of solvation effects, with COSMO-RS giving more realistic amounts of the *anti,anti*-conformer in THF and DMSO.

Introduction

The hydrogen-bond donating properties of urea and thiourea functional groups are exploited throughout the chemical sciences. Ureas and thioureas are now commonplace organocatalysts,[1] used to promote and control reactivity through noncovalent binding, while they are also used extensively in anion-recognition and binding.[2] It is therefore unsurprising that chemists' attention has focussed on characterizing the conformational behavior of (thio)ureas across several media (gas-phase, solution, solid-state) using myriad experimental techniques. The catalytic roles of (thio)ureas as catalysts has also been the subject of intense computational interest.[3]

In 2018 Gouverneur and co-workers reported the development of a new axially-chiral catalyst, containing two urea functional groups, which binds fluoride anions in a tridentate fashion similar to that found in the fluorinase enzyme active site.[4] The asymmetric nucleophilic fluorination of episulfonium ions was demonstrated in dichloromethane with this catalyst using otherwise insoluble caesium fluoride. The ability of ureas to adopt multiple conformations underlies this catalyst's mode of action: classical molecular dynamics simulations and quantum chemical calculations have shown this binding mode requires one urea to adopt a *syn-anti* conformation, while the other is *anti,anti*-configured (**Figure 1**). This was somewhat unexpected, since crystallographic observations of the *syn*,*anti*-urea conformation are scarce,[5] and X-ray structures of other fluoride:diarylurea complexes show exclusively *anti,anti*-urea conformations.[6] The tridentate binding mode has been in the solid-state following catalyst N-alkylation.

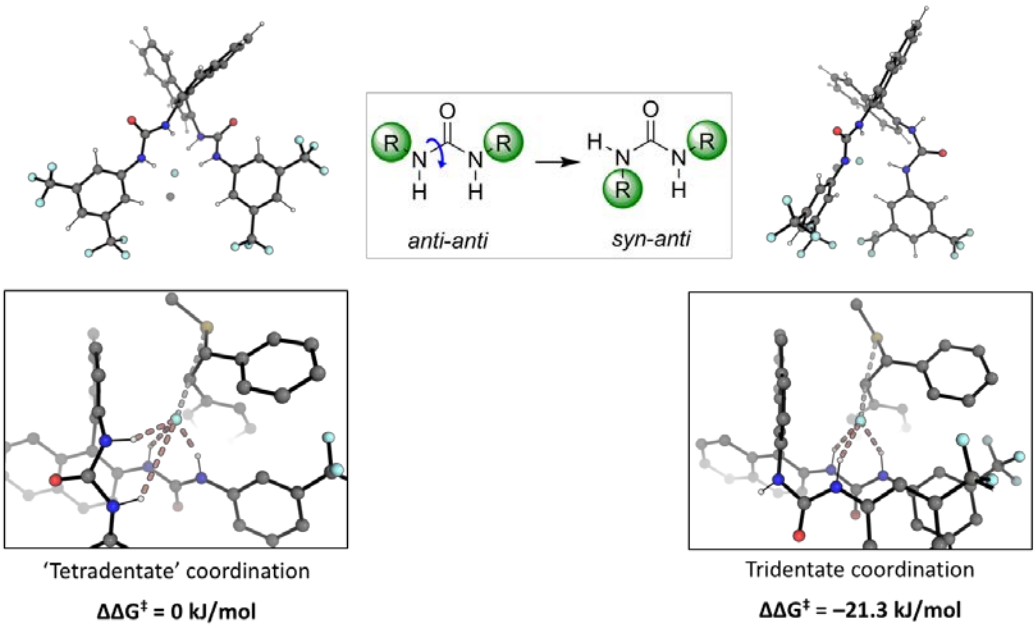


Figure 1. A urea *anti-syn* conformational change leading to distinct fluoride binding modes. The less common *syn,anti*-conformation results in a tridentate binding mode and smaller barriers (ωB97X-D3/(ma)-def2-TZVPP/COSMO(CH_2Cl_2)//M06-2X/def2-SVP(TZVPPD)/CPCM(CH_2Cl_2)) in nucleophilic fluorination.

Ureas infrequently observed in the aforementioned *syn,anti*-conformation. This is not the case, however, for thioureas, which exhibit more varied conformational behaviour. Schreiner, who has pioneered the development of thiourea-organocatalysts, found that diarylthioureas crystallize in both *syn,syn*- and *syn,anti*-conformations.[7] Jorgensen has reported the Gibbs energy difference at the G3B3 level switches from an *anti,anti*-preference by 1.09 kcal/mol for dimethylurea to a *syn,anti*-preference by 0.56 kcal/mol for dimethylthiourea.[8] ^{1}H and ^{13}C NMR studies of diphenylthiourea in $CDCl_3$ indicate the presence of both conformers in solution at room temperature, [9] while only the *anti,anti*-conformer of the di-3,5-bis(trifluoromethyl)phenylthiourea is observed in THF at room temperature. Suprisingly, dispersion-corrected density functional theory (DFT) calculations have been used to suggest that the *anti,anti*-conformer of diarylthioureas may be scarcely populated at all in dichloromethane.[10]

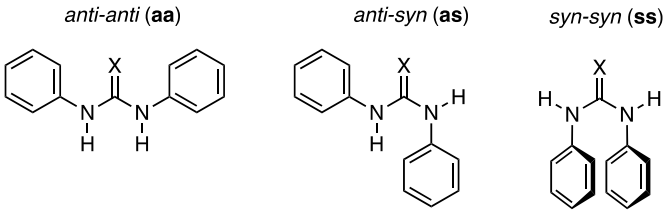


Figure 2. Conformers of diarylureas (X=O) and diarylthioureas (X=S).

We have investigated the conformational behavior of diarylureas and diarylthioureas using a combination of data-mining and quantum chemistry. Statistical analysis of a large number of structures is a useful way to infer the underlying conformational energetics: deviations from calculations on isolated molecules provides insight into intermolecular effects.[11] Using the Cambridge Structural Database (CSD)[12] Python API we have developed an automated tool to extract relevant structural information including the number of hydrogen-bonds formed by diaryl(thio)ureas in the solid state. We analyse these results in light of calculations on

diphenyl(thio)urea using different density functional approximations. We combine these two approaches to explore the intrinsic conformational preferences of (thio)ureas in different polarity solvents. We establish the underlying factors behind the qualitatively different behaviour exhibited by these two functional groups, which sees different populations of the three different conformations (**Figure 2**) in different conditions.

Results and Discussion

We first collected structural information from diaryl(thio)ureas found in the CSD. We developed a Python tool which allowed us to: (a) search for the presence of any organic fragment from its SMILES string; (b) automate the plotting of any pair of geometric variables (e.g. distances, angles, dihedrals); (c) analyse the noncovalent interactions (e.g. hydrogen-bonds) involving the organic fragment. This is generalizable to other structures. Searching for (thio)urea groups using the SMILES string 'X=C([NH]c1ccccc1)[NH]c1ccccc1' (where X is either O or S) and defining the two C-N-C-X dihedral angles results in the plots shown in **Figure 3**. Hexagonal clusters indicate regions (in terms of the two diehdrals) where several structures are found. The lower left corner corresponds to *anti,anti-* (aa) conformers, the upper right to *syn,syn-* (ss) conformers, and the upper left/lower right to *anti,syn-* (as) conformers.

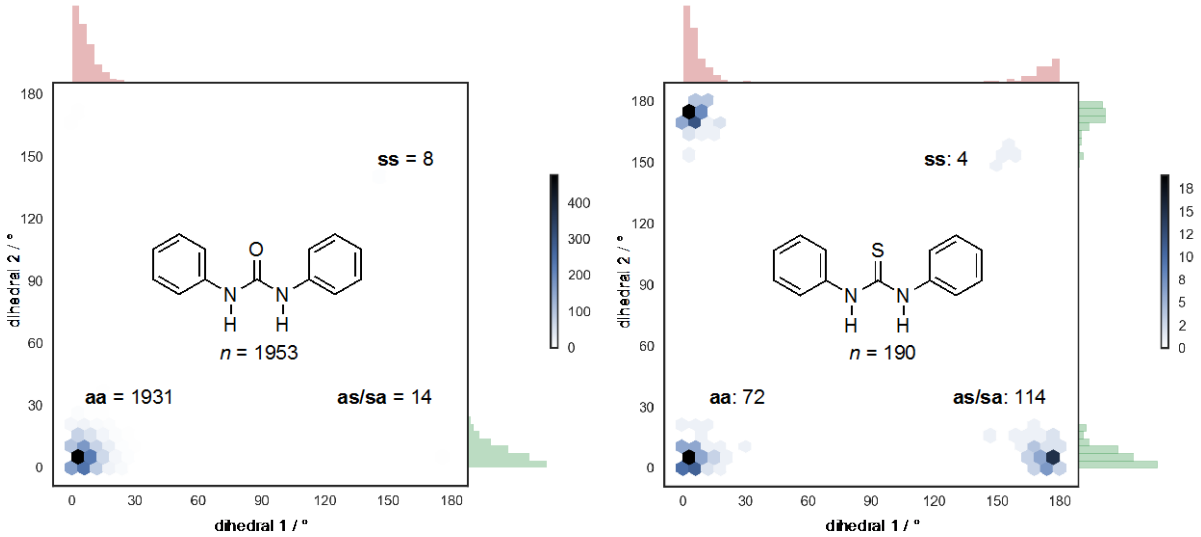
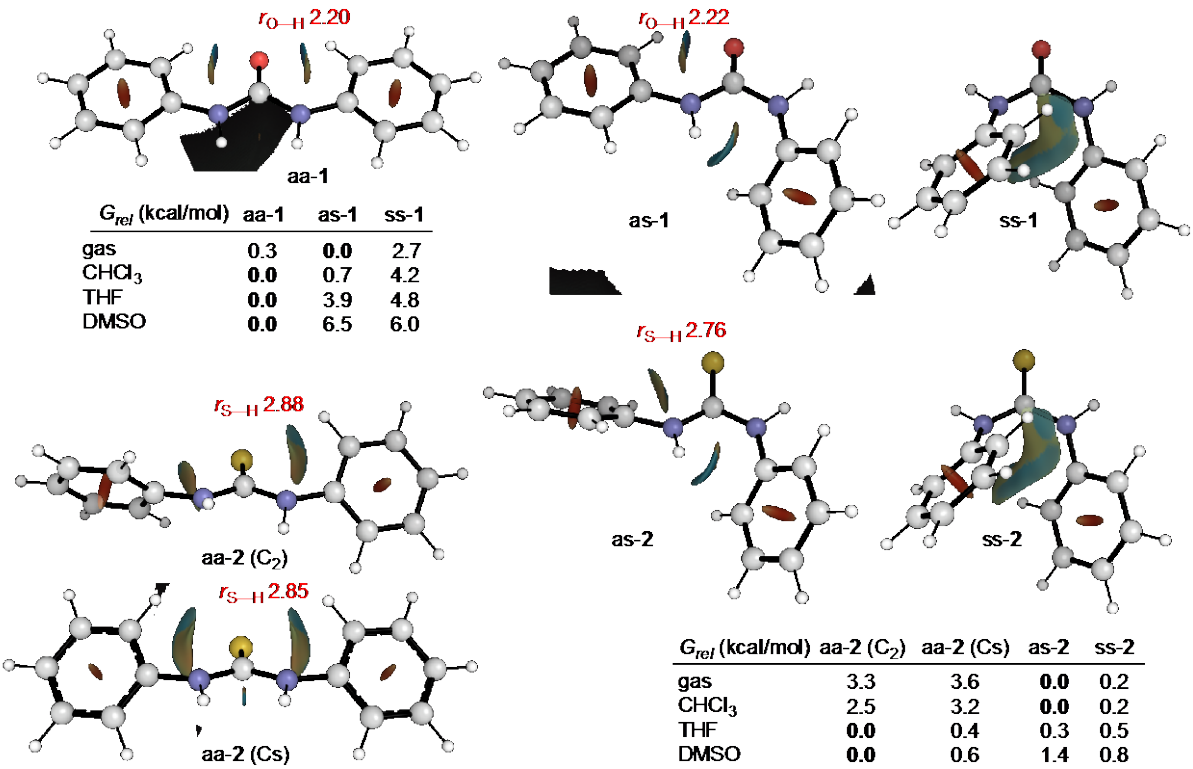


Figure 3. Bivariate Hexbin plots showing the counts and distribution for the two C-N-C=X dihedral angles of diphenylureas (LHS) and diphenylthioureas (RHS) found in the Cambridge Structural Database (CSD, version 5.39). Univariate histograms are also shown for each angle along the axes.

There are many more X-ray structures of diarylureas (1953) than thioureas (190). The overwhelming majority (99%) of these structures exist as aa-conformers (i.e. the C-N-C-X dihedral angles are clustered around 0°). The tiny proportion of as- and ss-conformers are not large enough to register on the plot. The rarity of these conformers is indicative of a sizable energetic penalty associated with them. In contrast, fewer than half of the diarylthioureas (38%) are found as aa-conformers. A more balanced distribution is visible in the plot, with most found in the as-form (60%). The ss-conformation (2%) makes a minor contribution and this cluster is now visible thanks to the smaller numbers elsewhere. As we shall discuss later, most of these structures form intermolecular N-H hydrogen-bonds, which influences conformation in addition to any "innate" preference of the isolated molecule. Nevertheless, the side-by-side comparison of **Figure 3** illustrates that thioureas show a much smaller preference for the aa-conformer.



G_{rel} (kcal/mol)	aa-1	as-1	ss-1
gas	0.3	**0.0**	2.7
$CHCl_3$	**0.0**	0.7	4.2
THF	**0.0**	3.9	4.8
DMSO	**0.0**	6.5	6.0

G_{rel} (kcal/mol)	aa-2 (C_2)	aa-2 (Cs)	as-2	ss-2
gas	3.3	3.6	**0.0**	0.2
$CHCl_3$	2.5	3.2	**0.0**	0.2
THF	**0.0**	0.4	0.3	0.5
DMSO	**0.0**	0.6	1.4	0.8

Figure 4. M06-2X-D3/def2TZVP structures of diphenylurea and diphenylthiourea conformers shown with NCI isosurface. Relative Gibbs energies shown for COSMO-RS treatment in different solvents. Intramolecular nonbonded contacts shown in Å.

To understand the origins of this difference, compare the optimized geometries of N,N'- diphenylurea (**1**) and N,N'- diphenylthiourea (**2**) in **Figure 4**. We found that structures were very similar at three different levels of theory (B3LYP-D3BJ, ωB97XD and M06-2X-D3) with the def2-TZVP basis set. The difference between **1** and **2** stems from the fact that the aa-form of **1** is planar, whereas, due to sulfur's larger size vs. oxygen, the aa-form of **2** is forced to adopt either of two bent conformations (according to whether the phenyl rings tilt in the same or opposite sense). Close interatomic contacts (less than van der Waals sum) and a blue-ish noncovalent interation index (NCI) isosurface point to the existence of intramolecular CH---X interactions in each case. The same out-of-plane rotation is present in the as-conformer of **2**, again caused by the larger steric volume of the sulfur atom. The ss-conformers of both **1** and **2** are essentially unchanged. These structures show that – qualitatively – differences in urea and thiourea conformational equilibria result from the relative destabilization of the planar aa-conformer of **1** caused by the sulfur atom.

How well do computed stabilities stack up against the quantitative trends extracted from the solid state? To address this we obtained Gibbs energy differences between the conformers above in the gas-phase and in different solvents. We show results obtained using a COSMO-RS (Conductor-like Screening MOdel for Real Solvents) treatment at the BP-TZVPD-FINE level (**Figure 4**).[13] Proportions of the three conformational states (aa, as, ss) were computed from the Boltzmann populations obtained from quasi-harmonic Gibbs energies at 298K in the gas-phase,[14] in chloroform ($CHCl_3$, ε=4.7), tetrahydrofuran (THF, ε =7.4) and dimethylsulfoxide (DMSO, ε =46.8). These solvents span a range of polarities and enable direct comparison against experimental NMR studies. The importance of dispersion-corrections has been

emphasised in studying (thio)urea conformations and so we show results from B3LYP-D3(BJ), M06-2X-D3 and ωB97XD calculations.[15] Reassuringly, relative conformer ratios are in broad agreement and there are no differences > 1 kcal/mol between the different methods. Additional tests with DLPNO-CCSD(T) and DSD-BLYP single point energy calculations also did not change the relative ordering of conformer stability.

Looking first at N,N'- diphenylurea (**1**), the aa-conformation is found to be scarce (<1%) under all conditions. Consistent with this relative instability, there are only a handful (8) of aa-diarylurea conformers in the CSD. Regarding the other two conformers, we compute that as, and ss-conformers are evenly balanced (with *as* slightly preferred) in the gas-phase, while the ss-conformer begins to dominate in $CHCl_3$, and does so completely in more polar solutions, as is observed in the solid state. These results compare favorably with observations from gas phase UV and mid-IR spectroscopy, showing that N,N'- diphenylurea preferably adopts an as-conformer over the ss-conformer.[16] In solution the ss-form starts to dominate and IR spectroscopy in DMSO shows an 88% population of the ss-conformer and a 12% population of the as-conformer.[17] The experimental trend towards increasing population of the ss-conformer in polar solvent is captured well by our calculations, and the relative scarcity of the aa-conformer is also consistent between all experiment and theory.

As outlined above, the conformational ensemble of N,N'-diphenylthiourea (**2**) is more complex and all three forms are computed to occur depending on the conditions. The aa-form of **2** appears in more polar, or Lewis basic solvents, whereas in the gas-phase the as-conformer is (slightly) preferred over the aa-conformer. Evidence for these equilibria is seen in the IR spectra of **2** in different chlorinated solvents.[18] NMR studies of diarylthioureas show the existence of a single species (the aa conformer) at room temperature in THF and toluene, with the as-form appearing at lower temperatures.[19] In the knowledge that the aa-conformer has been detected experimentally in a number of solvents, we found that COSMO-RS solvation performs rather better than other implicit solvation treatments, which predict a much smaller population of the aa-conformer. In solution, this is challenging equilibrium to describe computationally - the different conformers lie much closer in energy than for **1** and also have solvation energy terms differing by several kcal/mol. A balance between aa- and as-conformers is evident in the solid state and solution studies, which is captured in calculations in THF and DMSO (both Lewis basic solvents for which the COSMO-RS treatment of hydrogen-bonding plays an important role in stabilizing particularly the aa-form). The ss-conformer is present in four X-ray structures, however, experimental detection in solution is uncommon. For this reason, the computed population of ss-**2** does seem to be overexaggerated, although the results are considerably improved using COSMO-RS vs. SMD. The extent of $n_N - \pi^*_{C=X}$ delocalization in urea **1** is greatest in the planar, aa-conformer and contributes to this being the most stable form. This is supported through natural population analysis (NPA) (Figure S13). With $\pi^*_{C=S}$ being a better acceptor,[20] the magnitude of delocalization is greater in all conformers of thiourea **2** as compared to urea **1**, although in this case, the largest stabilization is conferred to the ss-conformer. The absence of planarity in the aa-form of **2** is such that delocalization no-longer favors this conformation.

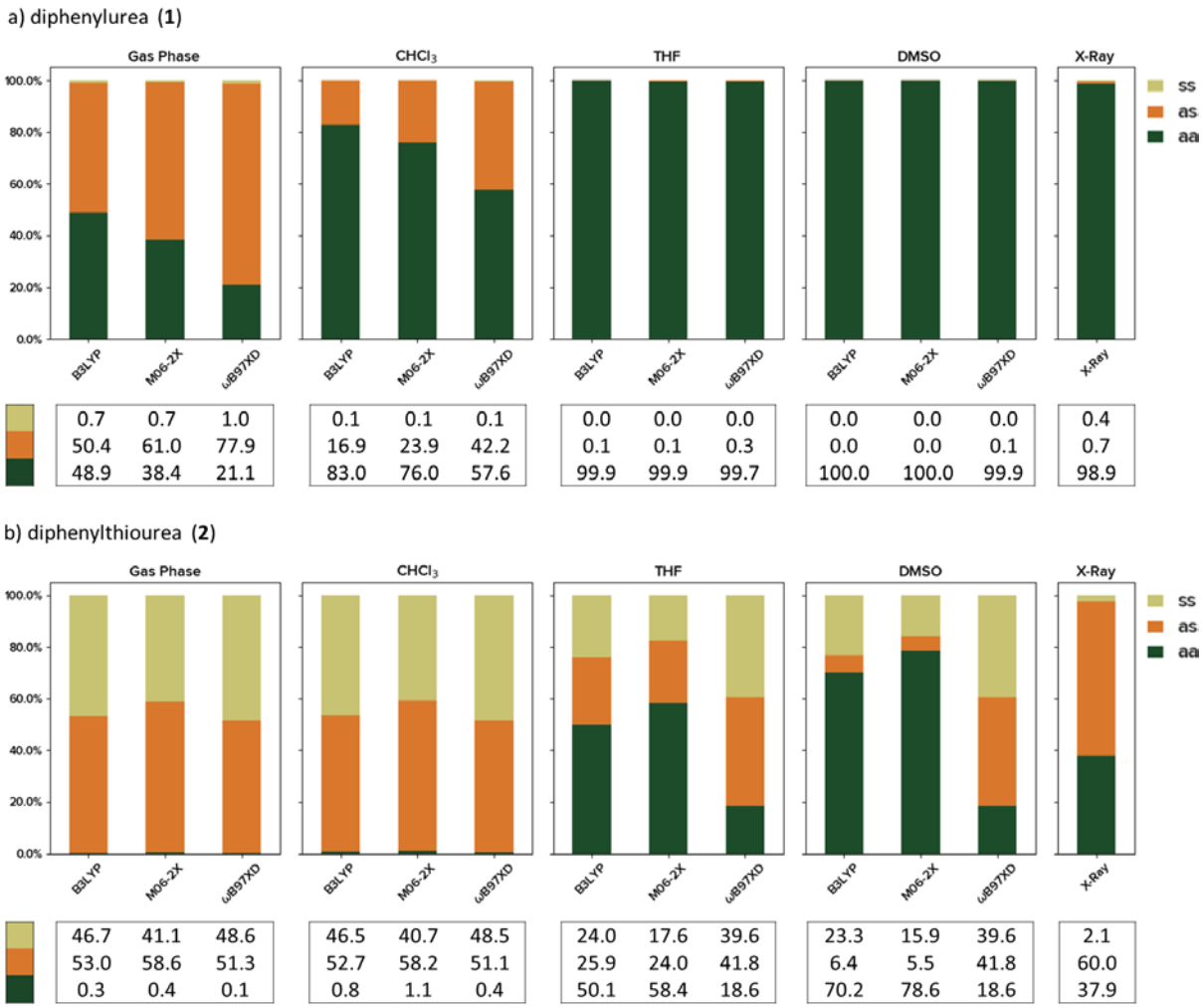


Figure 5. Computed equilibrium populations of the aa, as and ss-conformers of **1** and **2** at different levels of theory in different media (green = aa; orange = as; gold = aa). Levels of theory from left to right: B3LYP, M06-2X and ωB97X-D. Media from left to right: gas phase, $CHCl_3$, THF, DMSO and X-Ray. Values in the tables are shown in % and correspond to the functionals in the columns directly above.

To gain further understanding of how the solid-state results relate to our model, we used our Python tool to sort each X-ray conformer according to how many hydrogen-bonds are donated by the two N-H groups. We separated each cluster according to whether it was involved in zero, one, or two or more interactions of this type. A bifurcated hydrogen-bond, in which either one N-H donates to two acceptors, or two N-Hs donate to a single acceptor would be counted twice in this scheme. The most striking result is that aa-conformers are almost always engaged in at least two H-bonds (mostly involving the two N-Hs coordinated to a single acceptor), whereas the as-conformation is more likely to donate a single N-H hydrogen bond. Most (thio)ureas form polar H-bonding interactions in the solid-state and so the conformational distribution resembles that observed in polar solvents, which also preferentially stabilizes the ss-conformation. This result emphasises that computational models in implicit solvent are inherently fraught with difficulty, given the degree to which N-H coordination (to solvent molecules or other ureas/thioureas) will play in role in stabilizing low-energy conformations. The small number of ss-conformers makes a meaningful analysis difficult of these structures.

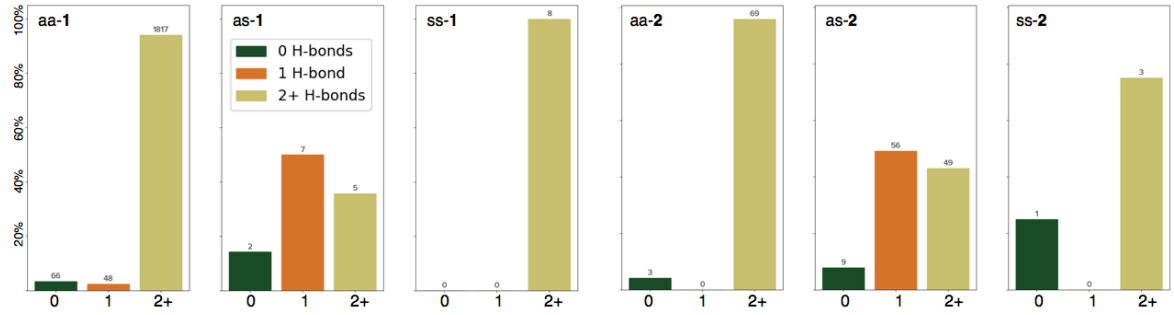


Figure 6. Number of N-H hydrogen-bonds formed by the ureas/thioureas in each conformation. These are grouped according to whether a given X-ray conformer forms zero, a single, or two or more hydrogen-bonds.

Conclusion

We developed a Python tool to allow the easy extraction of structural information from the Cambridge Structural Database using SMILES strings and automated clustering and visualization of selected geometric variables. This allowed us to compare the behavior of diarylureas and diarylthioureas: the former exhibits an overwhelming preference for the *anti,anti*-conformer, while the latter shows a mixture of *anti,anti*- and *anti,syn*-conformers. The origins of this difference were found to result from the destabilization of the *anti,anti*-thiourea, whereby the two aryl rings must rotate out of the plane to avoid a steric clash with the sulfur atom. This was confirmed by analysis of the NCI isosurface. The innate conformational preferences of N,N'-diphenylurea and N,N'-diphenylthiourea were studied under a variety of conditions using quantum chemistry. Apart from in the gas-phase, a preference for *anti,anti*-ureas was found and consistent with experiments this preference increased in more polar environments. Thiourea conformational behavior was found to show greater variability according to the medium. The proportion of *anti,anti*-conformer is predicted to increase with solvent polarity/Lewis basicity, while quantitative ratios showed much greater sensitivity to the computational treatment of solvation effects, with COSMO-RS giving more realistic amounts of the *anti,anti*-conformer in e.g. THF and DMSO.

Supporting Information

Computational methods and additional references. Absolute electronic energies, zero-point energies, enthalpies, entropies, Gibbs energies (a.u.) and Cartesian coordinates for all computed stationary points. Analysis using SMD solvation model. Python scripts for CSD analysis have been made available at DOI: 10.5281/zenodo.2045241

Corresponding Author

*robert.paton@colostate.edu

Acknowledgements

We acknowledge the EPSRC Centre for Doctoral Training in Synthesis for Biology and Medicine (EP/L015838/1) for a studentship to D. M. H. A., generously supported by AstraZeneca, Diamond Light Source, Defense Science and Technology Laboratory, Evotec, GlaxoSmithKline, Janssen, Novartis, Pfizer, Syngenta, Takeda, UCB and Vertex. R. S. P. acknowledges funding from Colorado State University. We used the Dirac cluster at Oxford supported by the EPSRC Centre for Doctoral training for Theory and Modelling in Chemical Sciences (EP/L015722/1); the RMACC Summit supercomputer supported by the National Science Foundation (ACI-

1532235 and ACI-1532236), the University of Colorado Boulder and Colorado State University; the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation TG-CHE180056. XSEDE is supported by the National Science Foundation (ACI-1548562).[21]

References

[1] (a) Schreiner, P. R. *Chem. Soc. Rev.* **2003**, *32*, 289–296; (b) Doyle, A.G.; Jacobsen, E.N. *Chem. Rev.* **2007**, *107*, 5713–5743.

[2] Jose, D. A.; Kumar, D. K.; Ganguly, B.; Das, A. *Org. Lett.*, **2004**, *6*, 3445–3448.

[3] (a) Cheong, P.H.Y.; Legault, C.Y.; Um, J.M.; Çelebi-Ölçüm, N.; Houk, K.N. *Chem. Rev.* **2011**, *111*, 5042–5137; (b) Hamza, A.; Schubert, G.; Soós, T.; Pápai, I. *J. Am. Chem. Soc.* **2006**, *128*, 13151–13160; (c) Gammack-Yamaguta, A. D.; Datta, S.; Jackson, K. E.; Stegbauer, L.; Paton, R. S.; Dixon, D. J. *Angew. Chem. Int. Ed.* **2015**, *127*, 4981–4985. (b) Madarász, A.; Dósa, Z.; Varga, S.; Soós, T.; Csámpai, A.; Pápai, I. *ACS Catal.*, **2016**, *6*, 4379–4387; (c) Peng, Q.; Duarte, F.; Paton, R. S. *Chem. Soc. Rev*. **2016**, *45*, 6093−6107.

[4] Pupo, G.; Ibba, F.; Ascough, D. M. H.; Vicini, A. C.; Ricci, P.; Christensen, K.; Morphy, J. R.; Brown, J. M.; Paton, R. S.; Gouverneur, V. *Science* **2018**, *360*, 638–642.

[5] Qureshi, N.; Yufit, D. S.; Steed, K. M.; Howard, J. A. K.; Steed, J. W. *CrystEngComm* **2016**, *18*, 5333–5337.

[6] (a) Engle, K. M.; Pfeifer, L.; Pidgeon, G. W.; Giuffredi, G. T.; Thompson, A. L.; Paton, R. S.; Brown, J. M.; Gouverneur, V. *Chem. Sci.* **2015**, *6*, 5293-5302; (b) Pfeifer, L.; Engle, K. M.; Pidgeon, G. W.; Sparkes, H. A.; Thompson, A. L.; Brown, J. M.; Gouverneur, V. *J. Am. Chem. Soc*. **2016**, *138*, 13314–13325.

[7] Kotke, M.; Schreiner, P. R. *Tetrahedron* **2006**, *62*, 434–439.

[8] Terhorst, J. P.; Jorgensen, W. L. *J. Chem. Theory Comput*. **2010**, *6*, 2762–2769.

[9] Sudha, L. V.; Sathyanarayana, D. N.; Bharati, S. N. *Magn. Reson. Chem.* **1987**, *25*, 474–479.

[10] Supady, A.; Hecht, S.; Baldauf, C. *Org. Lett*. **2017**, *19*, 4199−4202.

[11] Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. *Chem. Sci.* **2018**, *9* 2655-2665.

[12] Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C.; *Acta Cryst. B* **2016**, *B72*, 171–179.

[13] (a) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224–2235; (b) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. *J. Phys. Chem. A.* **1998**, 102, 5074–5085.

[14] (a) S. Grimme, *Chem. Eur. J.*, **2012**, *18*, 9955; (b) I. Funes-Ardoiz and R. S. Paton, **2016** GoodVibes v2.0.2. http://doi.org/10.5281/zenodo.595246. Vibrational scaling factors were automatically applied using data from the Truhlar group.

[15] DFT calculations used *Gaussian 16*, Revision A.03, Frisch, M. J. *et al.* Gaussian, Inc., Wallingford CT, **2016**. The Supporting Information contains a more detailed discussion of the computational methods used.

[16] (a) Emery, R.; Macleod, N. A.; Snoek, L. C.; Simons, J. P. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2816-2820; (b) Badawi, H. M.; Förner, W. *Spectrochim Acta, Part A* **2012**, *95*, 435–441.

[17] Sudha, L. V.; Sathyanarayana, D. N.; Bharati, S. N. *Magn. Reson. Chem.* **1987**, *25*, 474–479.

[18] Galabov, B.; Vassilev, G.; Neykova N.; Galabov, A. *J. Mol. Struct.* **1978**, *44*, 15-21.

[19] Lippert, K.M.; Hof, K.; Gerbig, D.; Ley, D.; Hausmann, H.; Guenther, S.; Schreiner, P. R. *Eur. J. Org. Chem.* **2012**, 5919–5927.

[20] Jackson, K. E.; Mortimer, C. L.; Odell, B.; McKenna, J. M.; Claridge, T. D. W.; Paton, R. S.; Hodgson, D. M. *J. Org. Chem.* **2015**, *80,* 9838–9846

[21] Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. *Comput. Sci. Eng.* **2014**, *16*, 62.