## Supplementary Material

**CSD Search Definitions**

Distributions in this publication were generated using ConQuest on the November 2013 release of the CSD system. Search results were further refined, and descriptors were extracted using python scripts written using the CSD Python Application Programming Interface (scripts have been included as part of the supplementary material). Structures were defined as organic if the entry contained only the elements C, Si, H, O, N, S, Se, P, Cl, F, Br, I. We note that the analysis omits common group I and group II salts.

**Definition of degrees of freedom**

The choice of definition was to reflect the typical number of degrees of freedom that direct space algorithms would face in solution, as typically such methods do not vary model bond lengths or angles, but do allow a molecule to freely translate and rotate in space while also varying molecular conformations.

A structure’s number of degrees of freedom was approximated as below in equation 1:

$dof= \sum\_{i}^{N}\left\{R\_{i}+ \left\{\begin{array}{c}3 if i is monoatomic\\5 if i is diatomic\\6 otherwise\end{array}\right\}\right\}$ (1)

Where Ri is the number of rotatable bonds in the ith molecular component in the structure of interest, the structure containing N components. In this analysis we treat terminal groups, such as methyl groups, hydroxyl groups and primary amines as rotatable. Usually hydrogen positions in such groups in structure solution would be ignored.

We note that the number of degrees of freedom is over-estimated by one for strictly linear molecular components (such as thiocyanates), but, given that the number of such molecules is relatively small in the CSD, this approximation has no significant effect on the generated results