Supporting Information

to the paper

Halogen bonding *versus* hydrogen bonding: what does the Cambridge Structural Database reveal?

by Tiddo J. Mooibroek and Patrick Gamez

Table of Contents

CSD searches

Search for potential C_6XY_5 host molecules and selection of the largest datasets. First, a dataset of 157,209 CIFs including C_6XY_5 aromatic rings (**X** = any atom; **Y** = H or Hlg) was retrieved from the CSD (version 5.33, May 2012). The data were exported as a .csv file for further analysis with Excel, producing a file of 168,766 hits (some CIFs contain more than one type of aromatic ring). These data were organized by type of C_6XY_5 ring, as shown in Table S1. The sub-dataset for **X** = C is significantly larger than that for **X** ≠ C (Table S1, see entries 1 and 2). Most of the C_6XY_5 rings are non-halogenated phenyl rings (Table S1, entry 3). Actually, 20,297 hits may involve halogenated aryl rings (entry 4). For the mono halogenated (Table S1, entries 5–7) and poly halogenated (Table S1, entries 8–12) rings, the series with most CSD hits is represented by the *p*-C_6XH_4 ring (Table S1, entry 7); therefore, this *p*-C_6XH_4 series was analysed thoroughly for potential intermolecular interactions.

Table S1. Number of CIFs found in the CSD for aromatic C_6XY_5 moieties.

Entry	**C_6XY_5 ring**[a]	**Halogen (Y, attached to the aromatic ring)**			
		F	**Cl**	**Br**	**I**
1	X ≠ C; Y_5 = H or Hlg	1,520			
2	X = C; Y_5 = H or Hlg	167,246			
3	X = C; Y_5 = H_5	146,949			
4	X = C; Y_5 ≠ H_5	20,297			
5	X = C; Y = *o*-Hlg; Y_4 = H_4	598	1,313	470	228
6	X = C; Y = *m*-Hlg; Y_4 = H_4	197	480	257	74
7	X = C; Y = *p*-Hlg; Y_4 = H_4	**2,291**	**4,685**	**3,744**	**414**
8	X = C; Y_2 = di-*o*-Hlg; Y_3 = H_3	352	555	12	0
9	X = C; Y_2 = di-*m*-Hlg; Y_3 = H_3	178			
10	X = C; Y_2 = *o,p*-Hlg; Y_3 = H_3	234	476	0	0
11	X = C; Y_5 = Hlg_5	2,898	74	2	0
12	X = C; Y_5 = other	765			

[a] *o*, *m* and *p* stand for *ortho*, *meta* and *para* relative to the **X** atom.

Table S2. Number of hits for each molecule···ElR interacting pair investigated (where molecule is a *para*-functionalized phenyl entity or dichloromethane)[a]

Entry	molecule	Number of hits								
		Total	**F**	**Cl**	**Br**	**I**	**N**	**O**	**P**	**S**
1	*p*-F-phenyl	6,738	2,352	370	50	54	1,314	2,251	78	258
2	*p*-Cl-phenyl	22,791	1,377	3,389	1,531	36	4,732	9,991	119	1,248
3	*p*-Br-phenyl	10,765	694	413	1,377	334	2,018	5,230	48	630
4	*p*-I-phenyl	2,349	293	100	13	312	417	1,022	14	188
5	*p*-Me-phenyl	43,618	6,209	4,013	542	315	7,441	20,243	567	4,185
6	CH_2Cl_2[b]	91,386	18,991	20,466	1,052	780	15,972	24,524	5,397	3,910
7	CH_2Cl_2[c]	68,880	12,662	18,412	789	507	12,365	18,038	3,061	2,792

[a] The subsets for ElR = As, Se, Te or At are not shown because they contain few hits. [b] Subsets for Cl_2CH_2···ElR interactions as depicted in Figure 1b (HCH plane). [c] Subsets for CH_2Cl_2···ElR interactions as depicted in Figure 1c (ClCCl in plane).

Procedure for the construction of modelled interacting host molecules and the determination of the related volumes.

The volumes occupied by the interacting host molecules (namely the *para*-functionalized phenyl rings and dichloromethane) have been calculated using Autodesk® Inventor® Professional 2012, with the metric parameters listed below (which have been retrieved from the literature or the CSD).

- *Para-functionalized phenyl rings*: first it was assumed that the hexagonal C_6 ring (phenyl ring) was flat. Literature values[1] were used for the aromatic bond distances C–C = 1.40 Å and C–H = 1.09 Å. The C–Hlg and C–CH_3 bond distances were averaged from the data retrieved from the CSD (see Table below).

- *Dichloromethane*: mean C–H/Cl distances and H–C–H / Cl–C–Cl angles as well as the average angle between the CH_2 and CCl_2 planes were calculated from the CSD data (see Table below).

Metric parameters used to construct model host molecules (*para*-functionalized phenyl rings or dichloromethane):

		Entity considered (number of hits)						
		***p*-F-Ph (3,844)**	***p*-Cl-Ph (7,390)**	***p*-Br-Ph (5,187)**	***p*-I-Ph (1,002)**	***p*-CH_3-Ph (29,509)**	**Ph (n.r.[a])**	**CH_2Cl_2 (174,547)**
Distances (Å)	**C–Hlg/CH_3**	1.362 ± 0.021	1.744 ± 0.030	1.899 ± 0.018	2.099 ± 0.018	1.502 ± 0.043	1.09[1]	1.730 ± 0.073
	C–H	n.r.[a]						0.981 ± 0.028
Angles (°)	**H–C–H**	n.r.[a]						108.02 ± 3.18
	Cl–C–Cl							112.45 ± 6.08
	H_2C–CCl_2							89.66 ± 2.15

[a] non-relevant due to the low amount of CSD data retrieved.

- *van der Waals (vdW) radii*: the literature values used were H = 1.09 Å, C = 1.70 Å, F = 1.47 Å, Cl = 1.75 Å, Br = 1.85 Å and I = 1.98 Å.[2,3]

- *The steric hindrance* of the methyl substituent of the *p*-CH_3-Ph group was modelled as a spherical volume of radius 2.16 Å (obtained from the C–H bond distance of 1.07 Å in a CH_3 moiety[1] combined with the vdW radius of H,[3] *i.e.* 1.09 Å).

- *Generation of the modelled interacting molecules (in space-filling mode):* Individual atoms (space-filling) were first generated as parts (.ipt extension) using Autodesk® Inventor® Professional 2012, and then assembled in a .iam file according to the metric parameters mentioned above (*i.e.* for the *para*-functionalized phenyl rings and dichloromethane). Using the 'Shrickwrap' function of Autodesk® Inventor® Professional 2012, the atoms were merged to generate a new .ipt file that contained the modelled molecule generated.

- *Computation of the volume of the modelled molecules as a function of* ***α; determination of the 'free' volume around a modelled molecule within a spherical segment as a function of α:*** Spherical cones (truncated if necessary to fit in the spherical segment) with increasing cone angle (from 5 to 90° with increments of 5°) were generated using Autodesk® Inventor® Professional 2012 as .ipt file. These (truncated) spherical cones where then collected in one .iam file and their centre of gravity aligned. These bodies are thus all located within one half of the spherical segment; mirroring these bodies then yields the (truncated) spherical cones that occupy the other half of the spherical segment. The modelled host molecule was then also placed in the same .iam file in such a way that the (truncated) spherical cones and the modelled host were properly aligned (see for example Figure 2). The interfering volume between the modelled host and the (truncated) spherical cones was then obtained using the 'analyze interference' algorithm in Autodesk® Inventor® Professional 2012, from which the volume in-between two ***α*** values was determined (Table S3). Subtracting the volume of the modelled molecule in-between two ***α*** values (see Table S3) from the volume difference between the (truncated) spherical cones corresponding to these two ***α*** values then gives the 'free' volume as a function of ***α*** (within a spherical segment around a modelled host molecule).

Table S3. Volumes (in Å^3) of the modelled molecules used as a function of $\boldsymbol{\alpha}$

	Interacting molecule							
α_a–α_b (°)[a]	***p*-F-Ph**	***p*-Cl-Ph**	***p*-Br-Ph**	***p*-I-Ph**	***p*-CH_3-Ph**	**Ph**	**CH_2Cl_2**[b]	**CH_2Cl_2**[c]
0–5	0.996	0.996	0.996	0.996	0.996	0.996	0.039	0.070
5–10	2.981	2.981	2.981	2.981	2.981	2.981	0.117	0.251
10–15	4.944	4.944	4.944	4.944	4.944	4.944	0.195	0.497
15–20	6.825	6.856	6.841	6.797	6.852	6.642	0.271	0.796
20–25	7.415	7.425	7.420	7.320	7.498	7.255	0.360	1.166
25–30	7.388	7.721	7.660	7.721	7.511	7.087	0.458	1.553
30–35	7.288	7.447	7.448	7.494	7.342	7.102	0.559	1.959
35–40	6.693	6.506	6.483	6.255	6.483	6.732	0.659	2.363
40–45	5.003	5.046	4.909	5.293	5.133	5.183	0.752	2.718
45–50	4.523	4.378	4.204	4.045	4.225	4.197	0.833	3.032
50–55	3.667	3.682	3.472	3.298	3.948	3.825	0.899	3.265
55–60	3.572	2.919	2.733	2.332	3.466	3.428	0.948	3.410
60–65	2.951	2.105	1.847	1.555	3.053	3.112	0.983	3.449
65–70	2.194	1.435	1.206	1.310	2.749	2.649	1.056	3.372
70–75	1.513	0.949	1.103	1.353	2.219	2.263	1.239	3.212
75–80	0.747	0.955	1.308	1.385	1.828	1.590	1.247	2.983
80–85	0.575	0.972	1.148	1.377	1.830	0.994	1.340	2.673
85–90	0.579	0.978	1.155	1.417	1.884	0.469	1.380	2.337
90-95	0.578	0.975	1.152	1.413	1.832	0.356	1.465	2.013
95–100	0.572	0.964	1.139	1.397	1.826	0.284	1.566	1.666
100–105	0.561	0.946	1.117	1.370	1.814	0.233	1.658	1.435
105–110	0.545	0.920	1.087	1.333	1.794	0.224	1.771	1.305
110–115	0.526	0.887	1.048	1.285	1.757	0.217	1.912	1.018
115–120	0.503	0.848	1.002	1.228	1.690	0.207	2.026	0.989
120–125	0.475	0.802	0.948	1.162	1.604	0.196	2.144	0.919
125–130	0.445	0.750	0.886	1.086	1.508	0.185	2.336	0.932
130–135	0.410	0.692	0.818	1.003	1.400	0.171	2.363	0.859
135–140	0.373	0.629	0.744	0.912	1.281	0.157	2.457	0.708
140–145	0.333	0.562	0.664	0.813	1.154	0.141	2.319	0.684
145–150	0.290	0.490	0.578	0.709	1.018	0.124	0.000	0.544
150–155	0.245	0.414	0.489	0.599	0.875	0.106	3.519	0.461
155–160	0.199	0.335	0.396	0.485	0.724	0.089	1.166	0.360
160–165	0.150	0.253	0.299	0.367	0.568	0.069	0.795	0.271
165–170	0.101	0.170	0.201	0.246	0.410	0.050	0.497	0.195
170–175	0.051	0.085	0.101	0.124	0.247	0.030	0.251	0.117
175–180	0.006	0.011	0.013	0.015	0.082	0.010	0.070	0.039
Total:	76.217	79.030	80.540	83.421	95.530	74.298	41.650	53.621

[a] $\boldsymbol{\alpha}$ = C–**Y**–ElR, $\boldsymbol{\alpha}^{\mathbf{H}}$ = HCH–C–ElR, or $\boldsymbol{\alpha}^{\mathbf{Cl}}$ = ClCCl–C–ElR (see also Figure 1); [b] see Figures 3a and 3b top; [c] see Figures 3b bottom and 3c.

Refinement of [*p*-functionalized phenyl···ElR] and [CH_2Cl_2···ElR] datasets. The total amount of hits found in the CSD for each molecule···ElR interaction investigated (molecule = *para*-functionalized phenyl group or dichloromethane) are listed in Table S2. The datasets with sufficient amounts of hits were used to generate the ***P*** *vs*. $\boldsymbol{\alpha}^{\mathbf{(H/Cl)}}$ and ***F*** *vs*. ***D*** plots used in the present study (see **Experimental** section in main text for details).

*** For the *para*-functionalized phenyl rings** (Figure 1a), the datasets for *p*-Br-phenyl (Table S2, entry 3) and *p*-CH_3-phenyl (Table S2, entry 5) were refined further. The respective ***P*** *vs.* ***α*** plots are shown in Figures S1a-b and S2a. Figures S1c-d and S2b illustrate the ***F*** *vs*. ***D*** plots for respectively the *p*-CH_3-phenyl···ElR and *p*-Br-phenyl···ElR datasets.

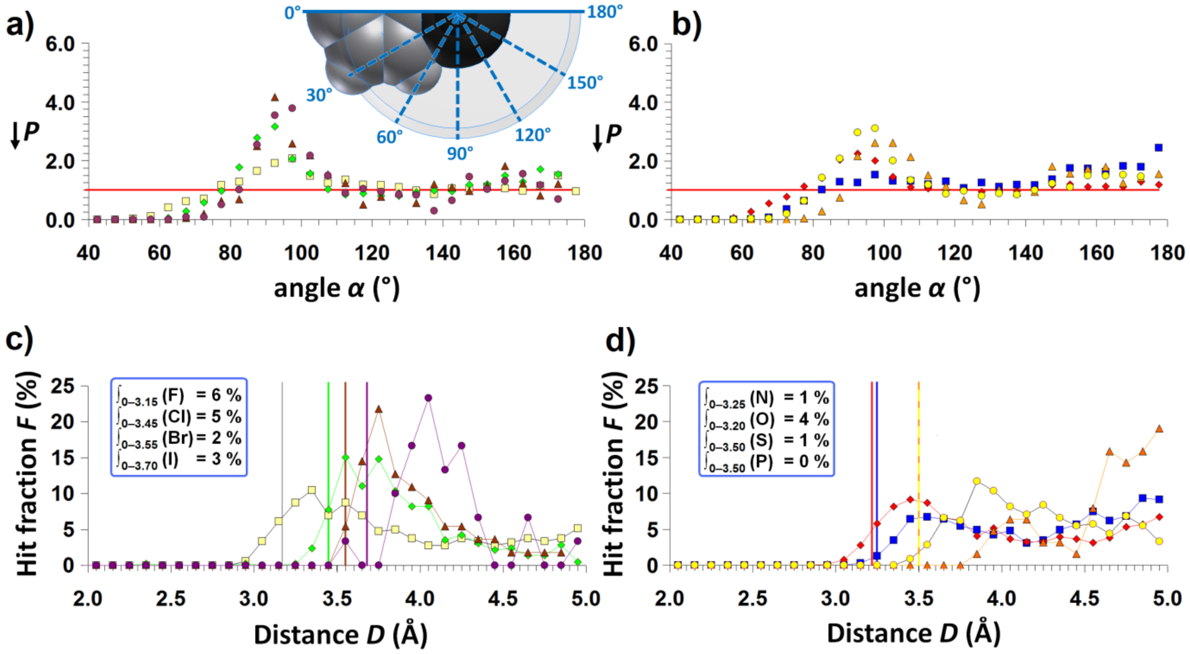

Figure S1. Volume-corrected probability ***P*** as a function of ***α*** for the *p*-CH_3-phenyl unit and a) common halogen atoms and b) non-halogen atoms; F (beige squares), Cl (green diamonds), Br (brown triangle), I (purple spheres), N (blue squares), O (red diamonds), S (yellow spheres), P (orange triangles). The inset figure in S1a shows a *p*-CH_3-phenyl ring where ***α*** angles with intervals of 30° are indicated (in blue) as an example; Hit fraction ***F*** as a function of ***D*** in the region ***α*** > 160° for c) common halogen atoms and d) non-halogen atoms. $\boldsymbol{N}_{\mathbf{tot}}$ = 505 (F), 425 (Cl), 55 (Br), 30 (I), 953 (N), 1,694 (O), 454 (S), 63 (P). The vertical lines delimit the respective Σ_{vdW} [C + F] distances and the inset values indicate the amount of hits (%) corresponding to vdW overlap.

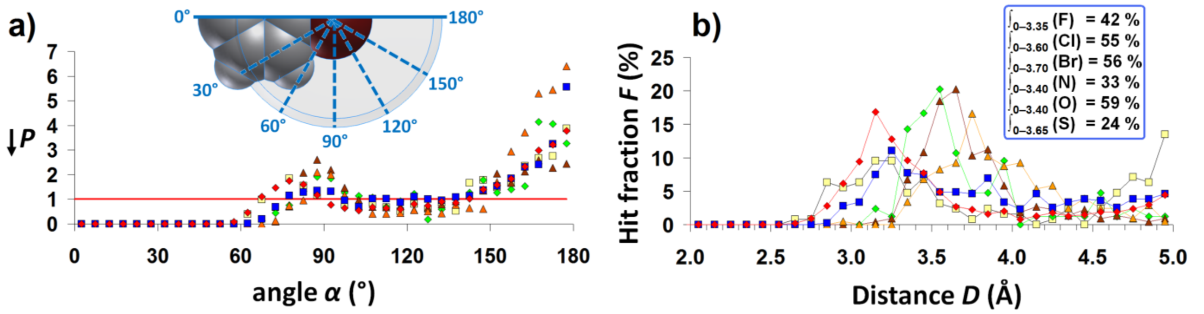

Figure S2. a) Volume-corrected probability ***P*** as a function of ***α*** for the *p*-Br-phenyl unit and F (beige squares), Cl (green diamonds), Br (brown triangle), N (blue squares), O (red diamonds), S (yellow spheres). The inset figure shows a *p*-Br-phenyl ring where ***α*** angles with intervals of 30° are indicated (in blue) as an example; b) Hit fraction ***F*** as a function of ***D*** in the region $\alpha > 160°$. N_{tot} = 694 (F), 413 (Cl), 1,377 (Br), 2,018 (N), 5,230 (O), 630 (S). The inset values indicate the amount of hits (%) corresponding to vdW overlap.

Since in all cases considered (namely for all aryl rings considered) the datasets for ElR = O contain numerous hits, ***P*** *vs*. ***α*** and ***F*** *vs*. ***D*** plots were generated for ElR = O and the different aryl rings investigated (see Figure S3).

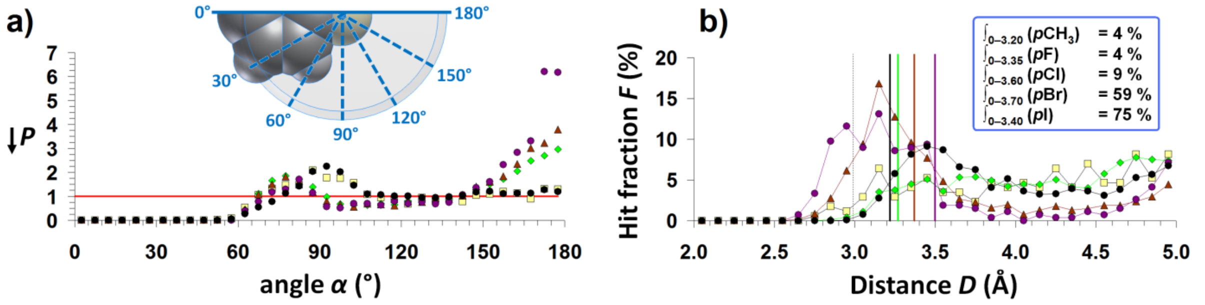

Figure S3. a) Volume-corrected probability ***P*** as a function of ***α*** for ElR = O and *p*-CH$_3$-phenyl (black spheres), *p*-F-phenyl (beige squares), *p*-Cl-phenyl (green diamonds), *p*-Br-phenyl (brown triangle) and *p*-I-phenyl (purple spheres). The inset figure shows a *p*-F-phenyl ring where ***α*** angles with intervals of 30° are indicated (in blue) as an example; b) Hit fraction ***F*** as a function of ***D*** in the region $\alpha > 160°$. N_{tot} = 1,694 (*p*-CH$_3$-phenyl), 2,251 (*p*-F-phenyl), 9,991 (*p*-Cl-phenyl), 5,230 (*p*-Br-phenyl) and 1,022 (*p*-I-phenyl). The vertical lines delimit the respective Σ_{vdW} [*p*-atom + O] distances and the inset values indicate the amount of hits (%) corresponding to vdW overlap.

*** For dichloromethane through its HCH plane** (see Figure 1b), the data given in Table S2, entry 6 were used to generate ***P*** *vs*. $\boldsymbol{\alpha}^{\mathbf{H}}$ plots for common halogen (Figure S4a) and non-halogen (Figure S4b) atoms. The corresponding ***F*** *vs*. $\boldsymbol{D}'$ plots for $\boldsymbol{\alpha}^{\mathbf{H}}$ = 30 – 80° are depicted in Figures S4c and S4d.

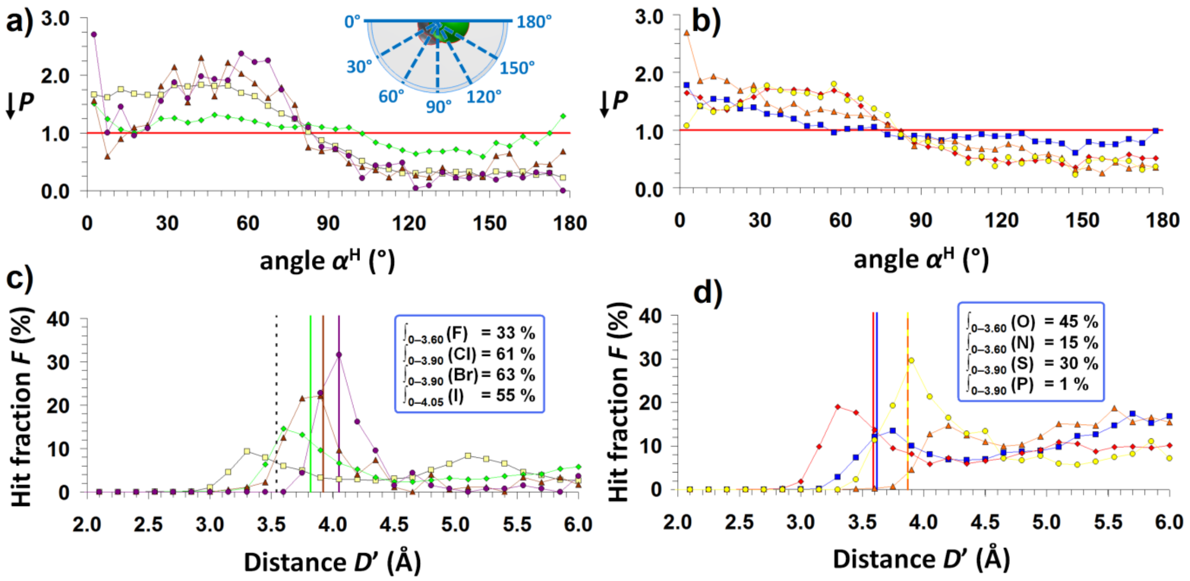

Figure S4. Volume-corrected probability ***P*** as a function of α^H (see Figure 1b) for a) common halogen atoms and b) non-halogen atoms; F (beige squares), Cl (green diamonds), Br (brown triangle), I (purple spheres), N (blue squares), O (red diamonds), S (yellow spheres) and P (orange triangles). The inset figure shows a CH_2Cl_2 molecule where α^H angles with intervals of 30° are indicated (in blue) as an example; Hit fraction ***F*** as a function of ***D′*** for c) common halogen atoms and d) non-halogen atoms in the region α^H = 30–80°. N_{tot} = 4,345 (F), 3,175 (Cl), 176 (Br), 136 (I), 3,165 (N), 4,933 (O), 778 (S) and 1,350 (P). The vertical lines delimit the respective C–H + Σ_{vdW} [H + ElR] distances and the inset values indicate the amount of hits (%) corresponding to vdW overlap.

* **For dichloromethane through its ClCCl plane** (see Figure 1c), the data given in Table S2, entry 7 were used to generate ***F*** *vs*. ***D′*** plots for α^{Cl} > 135°, which are depicted in Figures S5a for common halogen atoms, and Figure S5b for non-halogen ones.

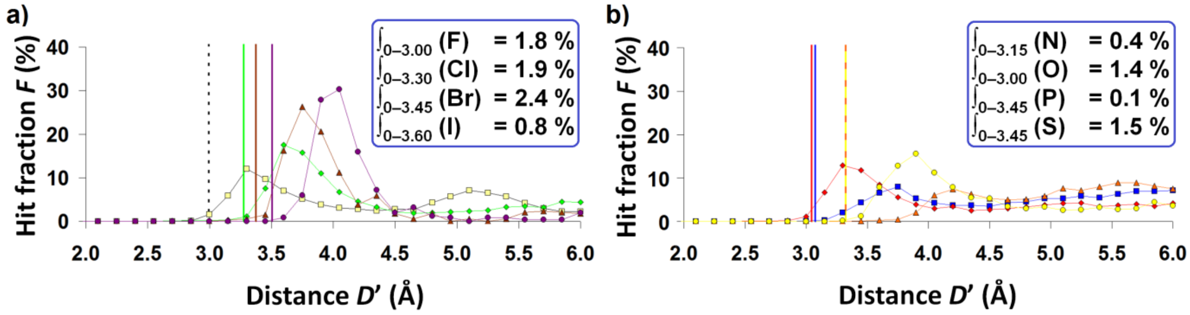

Figure S5. Hit fraction ***F*** as a function of ***D′*** for data found in the region α^{Cl} > 135° for a) common halogen atoms and b) non-halogen atoms; F (beige squares, N_{tot} = 6,647), Cl (green diamonds, N_{tot} = 5,698), Br (brown triangle, N_{tot} = 339), I (purple spheres, N_{tot} = 251), N (blue squares, N_{tot} = 5,043), O (red diamonds, N_{tot} = 8,229), S (yellow spheres, N_{tot} = 1,295), P (orange triangles, N_{tot} = 1,895). The vertical lines delimit the respective C–H + Σ_{vdW} [0.5 · H + ElR] distances and the inset values indicate the amount of hits (%) corresponding to vdW overlap.

References

(1) F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *J. Chem. Soc. Perkin Trans. II*, 1987, S1.

(2) A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441-451.

(3) R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384-7391.