SUPPLEMENTARY INFORMATION

Organic crystal hydrates: what are the important factors for formation?

Lourdes Infantes, László Fábián and W. D. Sam Motherwell

CONTENTS

1. Methodology - atom-group-codes

The methodology of how the data have been prepared and manipulated and, how the database CSDHBond has been created is described in previous publications and their supplementary sections in references 1 & 2.

[1] L. Infantes and S. Motherwell: The probable number of hydrogen-bonded contacts for chemical groups in organic crystal structures. *ChemComm.*, 2004, 1166-1167.
[2] L. Infantes and W. D. S. Motherwell: Hydrogen bond competition between chemical groups: new methodology and the Cambridge Structural Database. *Z. Kristallogr.*, 2005, **220**, 333-339.

In the Pluto program each donor and acceptor atom is identified as belonging to one of 83 atom types (e.g. OH, O double bond), and also to one of 113 chemical groups e.g. COOH, COO-, C=O keto, C-O-C, C(t3)-OH meaning a tertiary alcohol, having carbon with 3 covalent bonds. These chemical groups are scanned in a systematic order to assign atoms as far as possible to one of the groups. More general groups not covered earlier in the table of groups are assigned as more general such as say
B-OH. Rare chemical groups not in the table are given a group number zero. The chemical group connectivities are explicity coded in an input table to RPluto not shown here.

For example, there are three different atom-group codes for bromine atoms:

Br^- atom-group code=16000021 Br.0 Br(-)
Bromine with no covalent bonds, group number 21

X_3-Br atom-group code=16100109 Br.1 Others
Bromine has 1 covalent bond, connected to an atom which has three covalent bonds. No group assigned, 0.

C3-Br atom-group code 16100109 Br.1 Br(C3)
Bromine has 1 covalent bond, connected to carbon which has a total of 3 covalent bonds, group 109.

Some groups as originally coded in the group table were found to never to occur in our sample and others very infrequently. We present here the results for those groups with >100 occurrences in the sample.

Table 1. Functional groups with more than 100 observations in the CSDHBond database for 34770 CSD reference codes. Average values of parameters are given, as described in the main text. AtGroCod = Atom group code. AtType = atom type. GroupType = chemical group type. NDavg; NAavg; NDmax; NAmax; NDavg,In = NDavg,inter; NAavg,In = NAavg,inter; NDm,In = NDmax,inter; NAm,In = NAmax,inter; as defined in the text of the main paper. NObsAGC = number of observations of that chemical group in the total sample.

AtGroCod	AtType	GroupType	NDavg	NAavg	NDmax	NAmax	NDavg,In	NAavg,In	NDm,In	NAm,In	NObsAGC
8000000	As	Others									120
11000000	B	Others									2124
16000021	Br.0	Br(-)		1,832		5		1,832		5	607
16100000	Br.1	Others		0,095		3		0,065		3	431
16100109	Br.1	Br(C3)		0,032		1		0,009		1	1196
21000020	Cl.0	Cl(-)		2,353		7		2,353		7	1940
21100000	Cl.1	Others		0,064		3		0,043		3	1128
21100016	Cl.1	HCCl3									465
21100017	Cl.1	H2CCl2		0,004		1		0,004		1	282
21100087	Cl.1	CCl3		0,028		1		0,007		1	435
21100107	Cl.1	Cl(C3)		0,013		2		0,004		1	3225
21400103	Cl.4	ClO4(-)									329
32100000	F.1	Others		0,163		3		0,157		3	1969
32100086	F.1	CF3		0,009		1		0,004		1	3237
32100105	F.1	F(C3)		0,036		1		0,02		1	2173
43000022	I.0	I(-)		0,931		4		0,931		4	202
43100000	I.1	Others		0,101		2		0,093		2	258
43100111	I.1	I(C3)		0,047		1		0,005		1	190
43200000	I.2	Others		0,127		2		0,025		1	118
56000001	N.PL3	NO3(-)		0,008		1		0,008		1	250
56000055	N.PL3	NO2		0,005		1		0,001		1	3654
56000089	N.PL3	N(T3)		0,008		1		0,001		1	4238
56101099	N.1.1	CN		0,375		3		0,373		3	1909
56102060	N.1.2	N(-)-N(+)=N		0,015		1		0,007		1	134
56200057	N.2	C=N-NH		0,292		2		0,248		2	1515
56200058	N.2	C=N-OH		0,568		2		0,535		2	613
56200060	N.2	N(-)-N(+)=N		0,019		1		0,015		1	268
56200067	N.2	C-N=N-		0,362		3		0,282		3	1282
56200090	N.2	N(T2)		0,544		3		0,363		2	6837
56201000	N…1	Others	0,312	0,37	2	2	0,217	0,326	1	2	138
56300089	N.3	N(T3)		0,335		3		0,206		1	3488
56301083	N.3.1H	C(T4)NHC(T4)	0,325	0,278	2	1	0,245	0,21	2	1	1212
56301084	N.3.1H	C(T4)NHC(T3)	0,981	0,006	2	1	0,901		2		161
56301088	N.3.1H	NH	0,277	0,027	3	1	0,257	0,027	3	1	548
56302000	N.3.2H	Others	0,917	0,298	3	1	0,882	0,295	3	1	363
56302078	N.3.2H	H2NCHR2	0,523	0,636	3	1	0,485	0,621	2	1	132
56302082	N.3.2H	H2NCR2	1,915	0,003	4	1	1,866	0,003	4	1	305
56400000	N.4	Others		0,007		1					535
56401076	N.4.1H	NHR3(+)	1,049	0,001	3	1	0,887	0,001	3	1	1525
56402075	N.4.2H	NH2R2(+)	2,118	0,001	5	1	2,037	0,001	4	1	1479
56403074	N.4.3H	NH3(+)	3,243	0,003	7	2	3,196	0,003	7	1	2482
56404011	N.4.4H	NH4(+)	4,345	0,015	8	1	4,345	0,015	8	1	203
56501085	N.PY.1H	C(T3)NHC(T3)	1,13		3		1,083		2		1000
56502067	N.AR.2	C-N=N-		0,462		2		0,441		2	247
56502090	N.AR.2	N(T2)		0,538		2		0,472		2	4849

56503089	N…3	N(T3)		0,022		1		0,004		1	558
56600089	N.AM	N(T3)		0,003		1		0,001		1	5507
56601026	N.AM.1H	RHN-CO-NHR	0,953	0,002	2	1	0,857	0,001	2	1	875
56601028	N.AM.1H	RHN-CS-NHR	0,771		2		0,545		2		503
56601048	N.AM.1H	-O-CO-NH-	0,884	0,002	2	1	0,837	0,001	2	1	1152
56601050	N.AM.1H	CONH	0,894	0,003	2	1	0,74	0,001	2	1	8599
56601051	N.AM.1H	CSNH	0,823	0,002	2	1	0,695	0,002	1	1	486
56601057	N.AM.1H	C=N-NH	0,816	0,004	2	1	0,732	0,004	2	1	533
56602023	N.AM.2H	H2N-CO-NH2	1,857	0,003	3	1	1,857	0,003	3	1	286
56602024	N.AM.2H	H2N-CS-NH2	1,439		3		1,439		3		230
56602025	N.AM.2H	H2N-CO-NHR	1,63	0,022	4	1	1,507	0,022	4	1	138
56602027	N.AM.2H	H2N-CS-NHR	1,289		4		1,246		4		114
56602046	N.AM.2H	CONH2	1,78	0,004	3	1	1,666	0,004	3	1	919
56701035	N.PL3.1H	R2NHPO	0,937		2		0,898		2		127
56701057	N.PL3.1H	C=N-NH	0,979	0,006	2	1	0,651	0,005	2	1	964
56701084	N.PL3.1H	C(T4)NHC(T3)	0,812	0,007	2	1	0,586	0,005	2	1	2405
56701085	N.PL3.1H	C(T3)NHC(T3)	0,934	0,005	2	1	0,689	0,002	2	1	4439
56701088	N.PL3.1H	NH	0,73	0,007	2	1	0,639	0,005	2	1	1599
56702000	N.PL3.2H	Others	1,715	0,025	3	1	1,633	0,019	3	1	158
56702080	N.PL3.2H	H2NPh	1,189	0,152	4	1	0,877	0,149	3	1	1016
56702082	N.PL3.2H	H2NCR2	1,807	0,016	4	1	1,701	0,016	4	1	3372
64000094	O.2.CO	C=O		1,317		4		1,129		3	665
64200000	O.2	Others		0,743		3		0,73		3	522
64200103	O.2	ClO4(-)		0,373		3		0,373		3	987
64300066	O.3	COOR		0,271		1		0,217		1	129
64300097	O.3	-O-		0,224		2		0,189		2	1273
64302097	O.3.P	-O-		0,033		1		0,021		1	1282
64304097	O.3.N	-O-		0,113		2		0,094		2	564
64305097	O.3.O	-O-		0,078		1		0,063		1	206
64311048	O.3.2C	-O-CO-NH-		0,003		1		0,003		1	1152
64311066	O.3.2C	COOR		0,016		1		0,008		1	9311
64311097	O.3.2C	-O-		0,179		3		0,136		2	19434
64322097	O.3.2P	-O-		0,017		1		0,017		1	175
64390091	O.3.X.H	OH	0,92	0,453	2	2	0,856	0,441	2	2	735
64391012	O.3.C.H	HOCH3	0,939	0,844	2	3	0,939	0,844	2	3	494
64391013	O.3.C.H	HOCH2CH3	0,879	0,846	1	2	0,879	0,846	1	2	149
64391052	O.3.C.H	COOH	0,994	0,043	3	2	0,938	0,042	2	2	5722
64391068	O.3.C.H	HOCH2R	0,94	0,627	2	3	0,881	0,61	2	3	3572
64391069	O.3.C.H	HOCHR2	0,918	0,524	2	3	0,84	0,499	2	2	10481
64391070	O.3.C.H	HOCR3	0,87	0,284	2	3	0,702	0,258	2	3	5202
64391071	O.3.C.H	HOPh	0,985	0,215	3	2	0,602	0,186	3	2	6486
64391073	O.3.C.H	HOCR2	0,992	0,171	2	2	0,594	0,159	2	2	788
64392003	O.3.P.H	H2PO4(-)	0,988	0,273	1	2	0,988	0,273	1	2	242
64392030	O.3.P.H	RHPO3(-)	0,974	0,162	1	2	0,974	0,149	1	2	228
64392031	O.3.P.H	RH2PO3	0,978	0,065	1	1	0,97	0,065	1	1	232
64394058	O.3.N.H	C=N-OH	1,016	0,106	2	3	1,003	0,096	2	3	612
64394091	O.3.N.H	OH	1,036	0,156	3	2	0,943	0,156	3	2	333
64399010	O.3.2H	OH2	1,877	1,038	4	4	1,877	1,038	4	4	4616
64601053	O.2.CO2	COO(-)		1,474		4		1,41		4	5108
64602003	O.2.PO2	H2PO4(-)		2,161		3		2,161		3	242
64602030	O.2.PO2	RHPO3(-)		1,735		4		1,697		4	456
64602032	O.2.PO2	R2PO2(-)		1,154		4		1,144		4	508
64603005	O.2.SO2	HSO4(-)		1,054		3		1,054		3	111

64603006	O.2.SO2	SO4(2-)		1,712		4		1,712		4	344
64603038	O.2.SO2	RSO3(-)		1,012		4		0,992		4	1716
64603040	O.2.SO2	R2SO2		0,323		3		0,279		3	3594
64614001	O.2.NO3	NO3(-)		1,073		4		1,073		4	750
64624000	O.2.NO2R	Others		0,66		3		0,607		3	394
64624055	O.2.NO2R	NO2		0,199		3		0,127		3	7328
64702031	O.2.PO	RH2PO3		1,591		3		1,574		3	115
64702034	O.2.PO	C3PO		0,972		2		0,84		2	250
64702035	O.2.PO	R2NHPO		0,909		2		0,903		2	154
64702036	O.2.PO	R3PO		0,917		3		0,808		3	302
64703000	O.2.SO	Others		0,86		3		0,757		3	571
64704000	O.2.NO	Others		0,875		3		0,714		3	518
64710066	O.2.KETO	COOR		0,63		3		0,56		3	200
64710094	O.2.KETO	C=O		0,682		4		0,435		4	6871
64711066	O.2.ESTE	COOR		0,365		3		0,279		3	8799
64740046	O.2.AMID	CONH2		1,343		3		1,251		3	855
64740050	O.2.AMID	CONH		0,76		3		0,663		3	8614
64740093	O.2.AMID	HC=O		0,989		4		0,954		4	174
64740094	O.2.AMID	C=O		0,585		3		0,486		3	3657
64741048	O.2.AMES	-O-CO-NH-		0,598		2		0,472		2	1161
64741066	O.2.AMES	COOR		0,554		2		0,456		2	529
64744023	O.2.UREA	H2N-CO-NH2		2,413		4		2,413		4	143
64744025	O.2.UREA	H2N-CO-NHR		1,87		4		1,804		4	138
64744026	O.2.UREA	RHN-CO-NHR		1,121		3		1,056		3	639
64744050	O.2.UREA	CONH		0,582		2		0,551		2	920
64744094	O.2.UREA	C=O		0,4		2		0,364		2	305
64790093	O.2.ALDE	HC=O		0,675		2		0,379		2	277
64791052	O.2.ACID	COOH		0,717		4		0,669		4	5713
66300000	P…3	Others									236
66400000	P…4	Others									867
66400003	P…4	H2PO4(-)									121
66400030	P…4	RHPO3(-)									228
66400031	P…4	RH2PO3									116
66400032	P…4	R2PO2(-)									254
66400034	P…4	C3PO									250
66400035	P…4	R2NHPO									155
66400036	P…4	R3PO									321
66500000	P…5	Others									161
81200000	S.2.ria	Others		0,19		2		0,066		2	121
81200098	S.2.ria	-S-		0,024		1		0,006		1	168
81300000	S.2.rib	Others	0,015	0,022	1	1	0,015		1		134
81300098	S.2.rib	-S-		0,078		2		0,022		2	498
81311098	S.2.ric	-S-		0,04		2		0,013		1	3027
81391092	S.2.1H	SH	0,168	0,031	2	1	0,13	0,015	1	1	131
81600038	S.4.O2	RSO3(-)		0,025		1		0,014		1	570
81600040	S.4.O2	R2SO2		0,006		1					1791
81702000	S…5	Others		0,319		2		0,201		2	229
81710000	S…6	Others		0,089		2		0,026		1	191
81740051	S.1.am	CSNH		0,487		2		0,403		2	357
81740096	S.1.am	C=S		0,336		2		0,136		1	110
81744024	S.1.ure	H2N-CS-NH2		1,348		4		1,348		4	115
81744027	S.1.ure	H2N-CS-NHR		1,114		4		1,105		4	114
81744028	S.1.ure	RHN-CS-NHR		0,618		3		0,611		2	314
81744051	S.1.ure	CSNH		0,494		2		0,471		2	174

81790000	S…9	Others		0,03		1		0,012		1	165
84000000	Se…	Others									320
85000000	Si…	Others									1558
92000000	Te…	Others									125

2. CSD reference codes for the total sample (34770)
CSDHBondRefcodes.gcd

3. CSD refcodes for the hydrates (3258)
Hydrates.gcd

4. CSD refcodes for the water pattern DDA, DDAA, DD, DA (most common water donor-acceptor environments)
DDA.gcd
DDAA.gcd
DD.gcd
DA.gcd