

# Supporting Information

## Conformational Study of ( $8\alpha,8'\beta$ )-*bis*(Substituted Phenyl)-Lignano-9,9'-Lactones by Means of Combined Computational, Database Mining, NMR and Chemometric Approach

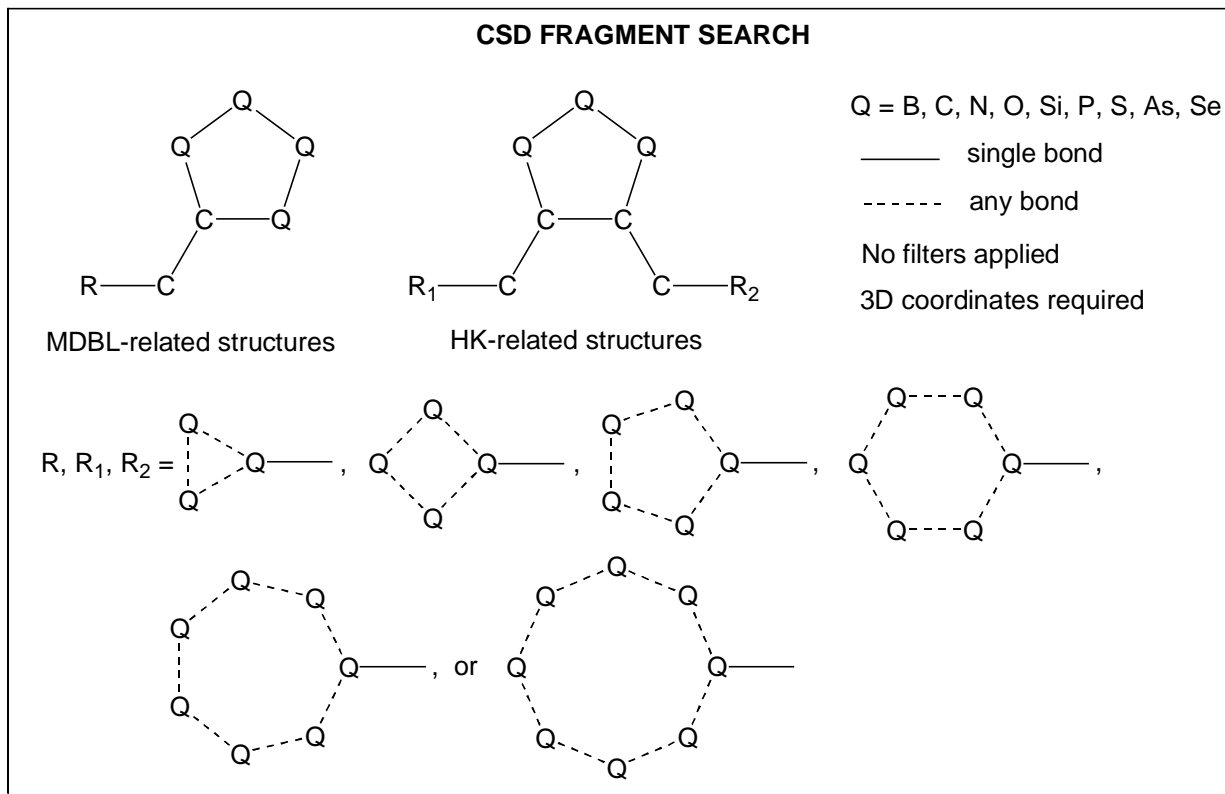
*Márcia M. C. Ferreira, Rudolf Kiralj, Paulo M. Donate, Rosangela da Silva, Sergio Albuquerque*

### CONTENTS with simple items descriptions

Scheme A. Data mining strategy for MDBL- and HK-related structures from the CSD .....	S3
Scheme B. Fragments used in the CSD search for 5-membered rings .....	S4
Figure A. Vista histograms for the rings from Scheme B .....	S4
Figure B. The three crystal structures as experimental reference for MDBL .....	S5
Table A. Experimental geometry of the structures from Figure B .....	S6
Table B. B3LYP basis set convergence for the MDBL geometry .....	S8
Figure C. The ten most similar structures to MDBL as retrieved from the CSD database .....	S10
Table C. Structural data for the ten MDBL-like compounds .....	S11
Description of structural descriptors in Table C .....	S12

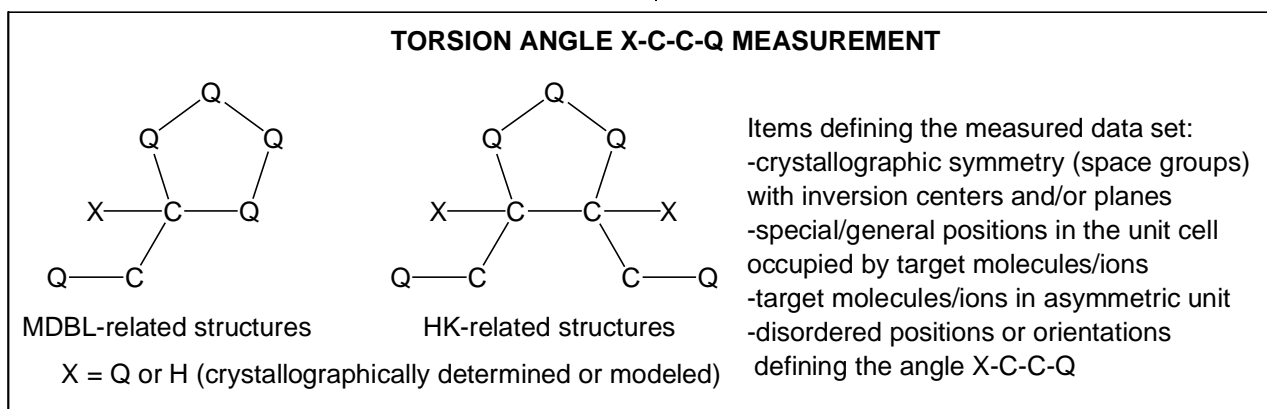
Figure D. Correlations of the structural descriptors with $\Delta\Theta$ for the MDBL-like structures .....	S14
Comments for Figure D .....	S14
Table D. Correlation coefficients between the descriptors from Figure D and $\Delta\Theta$ .....	S15
PLS analysis of data from Table C .....	S16
Figure E. PLS prediction of the PM3 rotation deviation $\Delta\Theta$ .....	S16
Figure F. The 27 HK-like structures retrieved from the CSD database .....	S17
Table E. Structural data for the 27 HK-like compounds .....	S18
Description of structural descriptors in Table E .....	S20
Figure G. Correlations of the structural descriptors with $\Delta\Theta$ for the HK-like structures .....	S20
Comments for Figure G .....	S21
Table F. Correlation coefficients between the descriptors from Figure D and $\Delta\Theta$ .....	S22
Table G. The deviation $\Delta\Theta$ for structure DIVJIQ from different computations .....	S23
Comments for Table G .....	S24
Figure H. Fragment defining the torsion angle CH-C-CH <sub>2</sub> -C .....	S25





**MANUAL STRUCTURE SELECTION**

<p><b>Eliminatory conditions:</b></p> <ul style="list-style-type: none"> <li>-central ring is not in spiro- or other ring structure</li> <li>-central ring has at most one double exocyclic bond</li> <li>-linker C is not in rings and is not bonded to metals</li> <li>-ring ambiguities avoided: only 1 and 2 X-C-C-Q torsion angles in MDBL- and HK-like structures, respectively, exist in retrieved structures</li> </ul>	<p><b>Other conditions:</b></p> <ul style="list-style-type: none"> <li>-R, R<sub>1</sub>, and R<sub>2</sub> may be bonded to metals and may form complicated rings and other structures</li> <li>-disordered structures are included</li> <li>-polymers and structures with errors are excluded</li> <li>-structures with no hydrogen atoms are included</li> <li>-all redeterminations are included</li> </ul>
---	---



Scheme A. Data mining strategy in obtaining torsion angles data for MDBL- and HK-related crystal structures from the Cambridge Structural Database (CSD).



Scheme B. Fragments that were used for definition of general  $\gamma$ -butyrolactone (left) and methylenedioxy (right) rings in the searches of Cambridge Structural Database (CSD), without any filtering. Bold are atoms defining the ring's least-squares plane, and the black balls are atoms on which the envelope is expected.

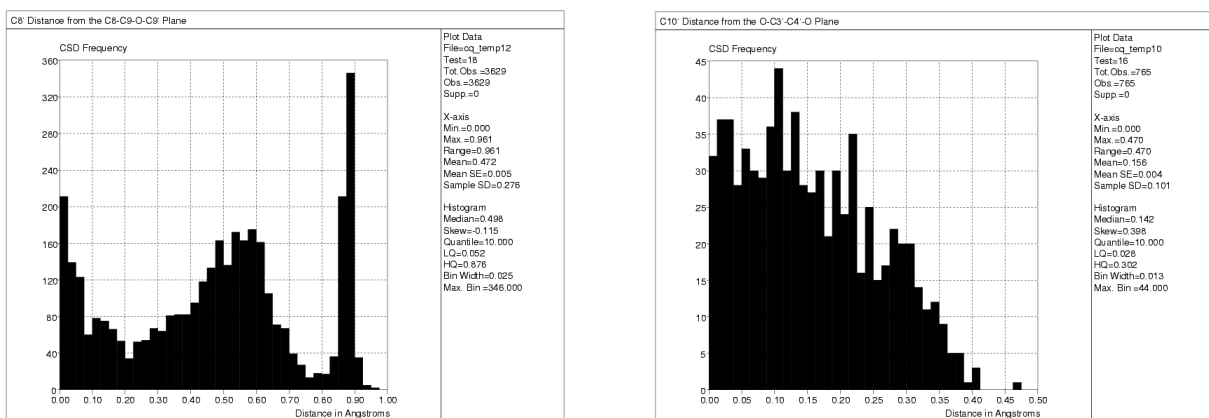


Figure A. Vista histograms representing the distance of the envelope atom from the least-squares plane of the other four atoms in the five-membered rings as defined in Scheme B. There are three maxima for  $\gamma$ -butyrolactone rings (left) and a continuum with decreasing frequency for methylenedioxy rings (right).

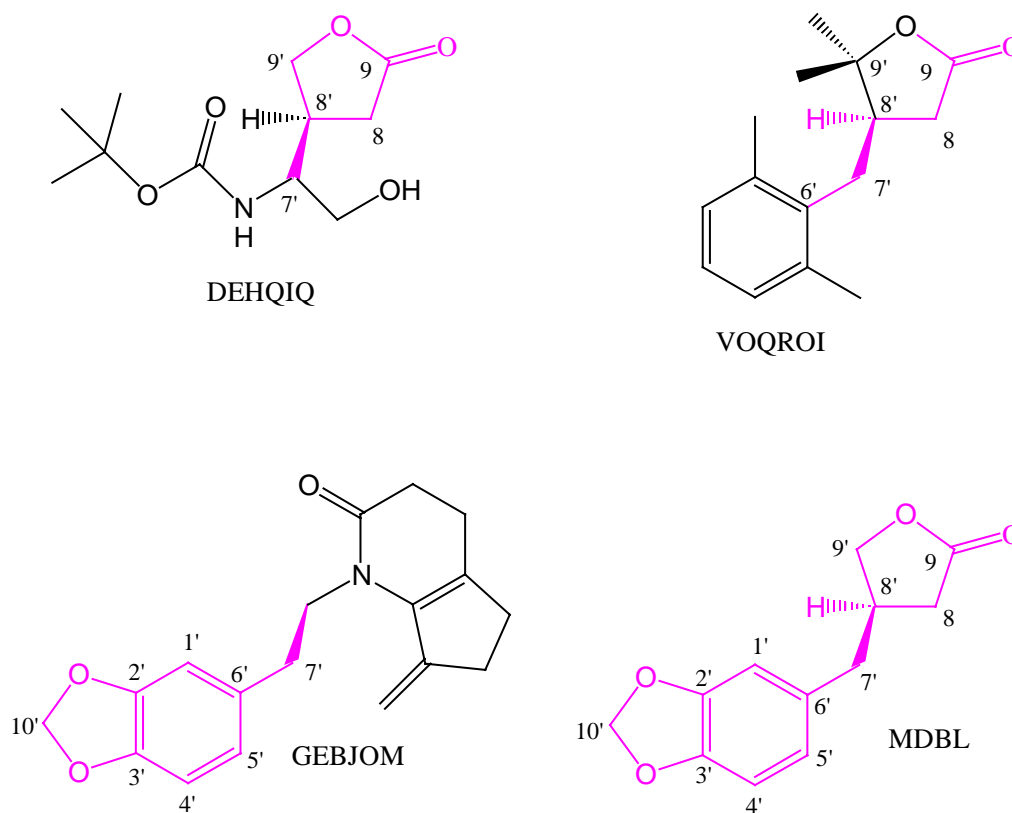


Figure B. Atomic numbering of MDBL applied to the three most relevant structures from the CSD. These three structures can be used instead of the inexistent crystal structure of MDBL, due to common parameter overlaps and similar structure determination quality and conditions, as shown below. Chemically and structurally identical fragments to those in MDBL are drawn pink.

CSD REFCODE	Temperature	R-factor	Cryst. System	Bond leng. exp.err.
DEHQIQ	room temp.	7.74%	orthorhombic	0.006 – 0.010 Å
GEBJOM	room temp.	5.60%	orthorhombic	0.006 – 0.010 Å
VOXROI	room temp.	6.12%	orthorhombic	0.006 – 0.010 Å

Table A. Experimental geometries used to define experimental values for MDBL to which results from various computational approaches should converge.

Parameter <sup>a</sup>	DEHQIQ <sup>b</sup>	VOQROI <sup>b</sup>	GEBJOM <sup>b</sup>	Recommended <sup>c</sup>	PM3 <sup>d</sup>
C9=O	1.203(6)	1.200(4)	-	<b>1.202(4)</b>	1.207
C9-O	1.358(7)	1.358(4)	-	<b>1.358(4)</b>	1.373
C9-C8	1.473(4)	1.487(4)	-	<b>1.480(4)</b>	1.510
C8-C8'	1.527(7)	1.516(4)	-	<b>1.522(4)</b>	1.534
C9'-O	1.467(9)	-	-	<b>1.467(9)</b>	1.428
C8'-C9'	1.535(9)	-	-	<b>1.535(9)</b>	1.547
C7'-C8'	1.526(7)	1.530(4)	1.528(6)	<b>1.528(3)</b>	1.527
C1'-C7'	-	1.509(3)	1.511(6)	<b>1.510(3)</b>	1.496
C1'-C6'	-	-	1.380(6)	<b>1.380(6)</b>	1.394
C1'-C2'	-	-	1.402(6)	<b>1.402(6)</b>	1.408
C2'-C3'	-	-	1.374(6)	<b>1.374(6)</b>	1.379
C3'-C4'	-	-	1.371(6)	<b>1.371(6)</b>	1.410
C4'-C5'	-	-	1.358(6)	<b>1.358(6)</b>	1.380
C5'-C6'	-	-	1.400(6)	<b>1.400(6)</b>	1.402
C3'-O	-	-	1.375(4)	<b>1.375(4)</b>	1.388
C4'-O	-	-	1.378(5)	<b>1.378(5)</b>	1.388
C10'-O(C3)	-	-	1.432(6)	<b>1.432(6)</b>	1.439
C10'-O(C4)	-	-	1.405(6)	<b>1.405(6)</b>	1.439
O-C9=O	120.3(6)	121.8(3)	-	<b>121.1(5)</b>	114.3
C8-C9=O	128.1(5)	129.5(3)	-	<b>128.8(3)</b>	133.1
O-C9-C8	111.6(4)	108.8(3)	-	<b>110.2(3)</b>	112.6
C9-C8-C8'	104.8(4)	103.1(2)	-	<b>104.0(2)</b>	104.1
C8-C8'-C9'	101.6(5)	102.7(2)	-	<b>102.2(3)</b>	104.6
C8-C8'-C7'	114.4(4)	117.2(2)	-	<b>115.8(2)</b>	113.0
C8'-C9'-O	105.7(5)	101.9(2)	-	<b>103.8(3)</b>	108.9
C7'-C8'-C9'	110.6(4)	116.6(2)	-	<b>113.6(2)</b>	111.3
C8'-C7'-C1'	-	111.5(2)	110.8(3)	<b>111.2(2)</b>	111.8
C7'-C1'-C6'	-	-	119.7(4)	<b>119.7(4)</b>	119.9
C7'-C1'-C2'	-	-	119.9(4)	<b>119.9(4)</b>	118.7
C1'-C6'-C5'	-	-	122.1(4)	<b>122.1(4)</b>	122.1
C6'-C1'-C2'	-	-	120.3(4)	<b>120.3(4)</b>	121.4
C1'-C2-C3'	-	-	116.8(2)	<b>116.8(2)</b>	116.3
C2'-C3'-C4'	-	-	121.8(4)	<b>121.8(4)</b>	122.1
C3'-C4'-C5'	-	-	122.9(4)	<b>122.9(4)</b>	121.9
C4'-C5'-C6'	-	-	116.0(4)	<b>116.0(4)</b>	116.2
C2'-C3'-O	-	-	127.1(4)	<b>127.1(4)</b>	128.3
C5'-C4'-O	-	-	127.1(4)	<b>127.1(4)</b>	128.5
O-C3'-C4'	-	-	110.2(4)	<b>110.2(4)</b>	109.6
O-C4'-C3'	-	-	110.0(4)	<b>110.0(4)</b>	109.6
C3'-O-C10'	-	-	104.3(3)	<b>104.3(3)</b>	106.1
C4'-O-C10'	-	-	105.1(4)	<b>105.1(4)</b>	106.1
O-C10'-O	-	-	104.3(3)	<b>104.3(3)</b>	108.7
O-C9'-C8'-C8	-27.2(6)	32.4(3)	-	<b>±29.8(3)</b>	-0.6
C9'-C8'-C8-C9	24.0(6)	-33.0(3)	-	<b>±28.5(3)</b>	0.7
C8'-C8-C9-O	-13.0(6)	21.5(3)	-	<b>±17.3(3)</b>	-0.6
C8-C9-O-C9'	-4.9(7)	-0.4(3)	-	<b>±2.7(4)</b>	0.2
C9-O-C9'-C8'	20.8(6)	-20.6(3)	-	<b>±20.7(3)</b>	0.3
H8'-C8'-C7'-C1' (Θ)	-	52.0	-	<b>52.0</b>	52.5
O=C9-O-C9'	176.6(6)	179.7(3)	-	<b>±178.2(3)</b>	-179.8
O=C9-C8-C8'	165.4(6)	-158.7(4)	-	<b>±162.1(4)</b>	179.4

$\Phi_{\text{CP}\gamma}$	243.1	107.0	-	<b>any multiple of 36°</b>	planar
$k_{\gamma}$	6.75	2.97	-	<b>close to an integer</b>	planar
$\Delta\text{C8}$	+0.430(5)	-0.534(2)		<b><math>\pm 0.482(3)</math></b>	+0.012
$\Phi_{\text{CPO}}$	-	-	325.3	<b>any multiple of 36°</b>	planar
$k_{\text{O}}$	-	-	9.04	<b>close to an integer</b>	planar
$\Delta\text{C10}'$	-	-	$\pm 0.127(6)$	<b><math>\pm 0.127(6)</math></b>	+0.002

<sup>a</sup>Parameters given in Å (angstrom) units are bond lengths and deviations of the envelope atoms C8 and C10' from the least-squares planes of the four atoms in the respective five-membered rings. Parameters given in (°) degrees are bond angles, torsion angles and Cremer-Pople puckering parameters for the  $\gamma$ -butyrolactone ring ( $\Phi_{\text{CP}\gamma}$ ) and the methylenedioxy ring ( $\Phi_{\text{CPO}}$ ). Without units are numbers  $k_{\gamma}$  and  $k_{\text{O}}$  obtained from general relationship  $\Phi = k \cdot 36^\circ$ . All parameters have experimental errors *i.e.* estimated standard deviations in brackets whenever possible. The positive and negative signs of deviations  $\Delta$  mean that the corresponding envelope atom lies above and below the plane of the other four atoms, respectively, in the  $\gamma$ -butyrolactone and methylenedioxy rings as defined in Figure B.

<sup>b</sup>The three MDBL-related crystal structures that were retrieved from the CSD, as defined in Figure B.

<sup>c</sup>Recommended values are made as averages whenever possible, otherwise are from the only crystal structure that possesses the MDBL's structural fragment of interest.

<sup>d</sup>Values obtained from PM3 geometry optimization of the MDBL geometry.



Table B. Detailed basis set convergence at the B3LYP level of the MDBL's molecular geometry.

Parameter <sup>a</sup>	STO-3G	3-21G*	6-31G*	6-31G**	6-31+G*	6-31+G**	6-311G*	6-311G**	6-311+G*	6-311+G**
C9=O	1.244	1.215	1.203	1.203	1.205	1.204	1.196	1.191	1.196	1.197
C9-O	1.440	1.406	1.367	1.367	1.366	1.365	1.366	1.372	1.365	1.363
C9-C8	1.579	1.536	1.525	1.525	1.523	1.521	1.522	1.525	1.519	1.522
C8-C8'	1.557	1.540	1.535	1.535	1.536	1.536	1.534	1.533	1.534	1.533
C9'-O	1.484	1.481	1.441	1.441	1.446	1.450	1.441	1.444	1.444	1.445
C8'-C9'	1.576	1.548	1.543	1.542	1.541	1.541	1.540	1.542	1.539	1.537
C7'-C8'	1.568	1.554	1.549	1.548	1.539	1.539	1.545	1.545	1.548	1.547
C1'-C7'	1.545	1.516	1.515	1.514	1.517	1.516	1.513	1.512	1.511	1.509
C1'-C6'	1.414	1.400	1.399	1.399	1.400	1.397	1.394	1.392	1.394	1.395
C1'-C2'	1.425	1.415	1.415	1.415	1.415	1.417	1.411	1.410	1.410	1.409
C2'-C3'	1.401	1.377	1.379	1.379	1.379	1.379	1.374	1.375	1.375	1.374
C3'-C4'	1.422	1.397	1.395	1.395	1.395	1.394	1.391	1.387	1.391	1.391
C4'-C5'	1.403	1.378	1.379	1.379	1.379	1.380	1.375	1.374	1.375	1.375
C5'-C6'	1.416	1.409	1.401	1.409	1.411	1.412	1.406	1.406	1.404	1.405
C3'-O	1.427	1.402	1.374	1.374	1.377	1.377	1.374	1.368	1.374	1.371
C4'-O	1.426	1.403	1.375	1.376	1.378	1.378	1.375	1.375	1.371	1.373
C10'-O(C3)	1.494	1.480	1.432	1.433	1.435	1.435	1.431	1.432	1.431	1.432
C10'-O(C4)	1.495	1.478	1.431	1.432	1.433	1.436	1.429	1.428	1.428	1.431
O-C9=O	120.7	121.7	122.3	122.3	122.0	122.1	122.5	122.2	122.5	122.3
C8-C9=O	129.3	129.8	128.6	128.7	128.6	128.3	128.5	128.2	128.8	128.9
O-C9-C8	110.0	108.5	109.1	109.1	109.4	109.6	109.0	109.1	108.8	108.8
C9-C8-C8'	104.0	103.7	103.9	103.9	103.5	103.7	104.4	104.8	104.1	103.8
C8-C8'-C9'	102.3	101.0	100.8	100.8	101.3	101.6	101.0	100.9	100.7	100.6
C8-C8'-C7'	115.1	111.4	113.0	1112.9	115.8	115.7	113.1	114.0	113.3	113.1
C8'-C9'-O	108.0	104.6	106.1	106.0	101.3	105.6	106.2	106.5	105.8	105.7
C7'-C8'-C9'	112.6	110.1	111.5	111.5	114.0	114.0	111.9	111.2	111.5	111.3
C8'-C7'-C1'	112.4	112.6	113.9	114.0	112.6	112.3	113.5	117.3	113.6	113.6
C7'-C1'-C6'	120.6	120.9	121.0	120.9	120.9	121.2	120.9	118.9	120.6	121.1
C7'-C1'-C2'	119.4	119.3	119.4	119.4	119.2	118.8	119.4	122.2	119.7	119.1
C1'-C6'-C5'	122.1	121.6	122.2	122.3	122.2	122.1	122.2	122.8	122.4	122.2
C6'-C1'-C2'	120.0	119.7	119.7	119.7	119.8	119.9	119.6	118.9	119.6	119.8
C1'-C2-C3'	117.9	118.0	117.6	117.6	117.4	117.4	117.7	118.0	117.7	117.5
C2'-C3'-C4'	121.5	121.8	122.2	122.2	122.3	122.2	122.1	122.3	121.9	122.2
C3'-C4'-C5'	121.2	121.4	121.6	121.6	121.7	121.7	121.6	121.1	121.8	121.6
C4'-C5'-C6'	117.3	117.4	116.8	116.8	116.6	116.8	116.8	116.9	116.6	116.8

C2'-C3'-O	127.0	127.7	128.2	128.1	128.2	128.3	128.5	128.0	128.7	128.2
C5'-C4'-O	127.4	128.1	128.8	128.8	128.9	128.9	129.0	129.2	128.7	128.9
O-C3'-C4'	111.5	110.5	109.6	109.6	109.5	109.6	109.4	109.7	109.4	109.5
O-C4'-C3'	111.5	110.5	109.6	109.6	109.4	109.4	109.4	109.7	109.4	109.5
C3'-O-C10'	102.9	105.9	105.5	105.5	104.9	104.9	105.0	104.8	104.8	105.0
C4'-O-C10'	102.9	105.9	105.5	105.5	105.0	104.9	105.0	104.8	104.9	105.0
O-C10'-O	111.3	107.1	108.0	108.0	107.3	107.3	107.4	107.8	107.3	107.3
O-C9'-C8'-C8	28.1	-35.1	-30.3	-30.3	29.9	29.5	-29.3	-30.4	-30.5	-31.5
C9'-C8'-C8-C9	-24.4	32.2	27.9	27.9	-28.0	-27.0	26.3	27.6	28.6	29.2
C8'-C8-C9-O	13.9	-18.6	-17.1	-16.9	17.6	16.1	-15.1	-16.4	-17.8	-18.0
C8-C9-O-C9'	3.9	-4.1	-2.7	-2.8	1.7	3.1	-4.1	-3.4	-2.2	-2.6
C9-O-C9'-C8'	-20.3	25.2	21.4	21.6	-20.5	-21.1	21.7	21.9	21.3	22.2
H8'-C8'-C7'-C1' ( $\Theta$ )	59.0	62.5	59.4	59.5	60.4	60.0	62.0	62.7	63.0	59.0
O=C9-O-C9'	-176.6	175.4	177.2	177.0	-178.2	-176.9	175.7	177.4	177.9	177.5
O=C9-C8-C8'	-165.6	162.0	163.1	163.3	-162.4	-163.9	165.1	162.7	162.2	162.0
$\Phi_{\text{CPY}}$	65.6	246.9	248.7	248.4	70.1	67.6	246.1	247.5	249.5	249.0
$k_{\gamma}$	1.82	6.86	6.91	6.90	1.95	1.88	6.84	6.88	6.93	6.92
$\Delta\text{C8}$	-0.441	+0.557	+0.488	+0.489	+0.486	+0.473	+0.466	+0.485	+0.495	+0.507
$\Phi_{\text{CPO}}$	planar	planar	323.8	324.0	288.0	323.7	323.6	323.2	322.9	324.3
$k_0$	planar	planar	8.99	9.00	8.00	8.99	8.99	8.98	8.97	9.01
$\Delta\text{C10}'$	+0.011	-0.018	-0.189	-0.197	-0.289	-0.286	-0.277	-0.282	-0.296	-0.281

<sup>a</sup>Parameters given in Å (angstrom) units are bond lengths and deviations of the envelope atoms C8 and C10' from the least-squares planes of the four atoms in the respective five-membered rings. Parameters given in ( $^{\circ}$ ) degrees are bond angles, torsion angles and Cremer-Pople puckering parameters for the  $\gamma$ -butyrolactone ring ( $\Phi_{\text{CPY}}$ ) and the methylenedioxy ring ( $\Phi_{\text{CPO}}$ ). Without units are numbers  $k_{\gamma}$  and  $k_0$  obtained from general relationship  $\Phi = k \cdot 36^{\circ}$ . The positive and negative signs of deviations  $\Delta$  mean that the corresponding envelope atom lies above and below the plane of the other four atoms, respectively, in the  $\gamma$ -butyrolactone and methylenedioxy rings as defined in Figure B.

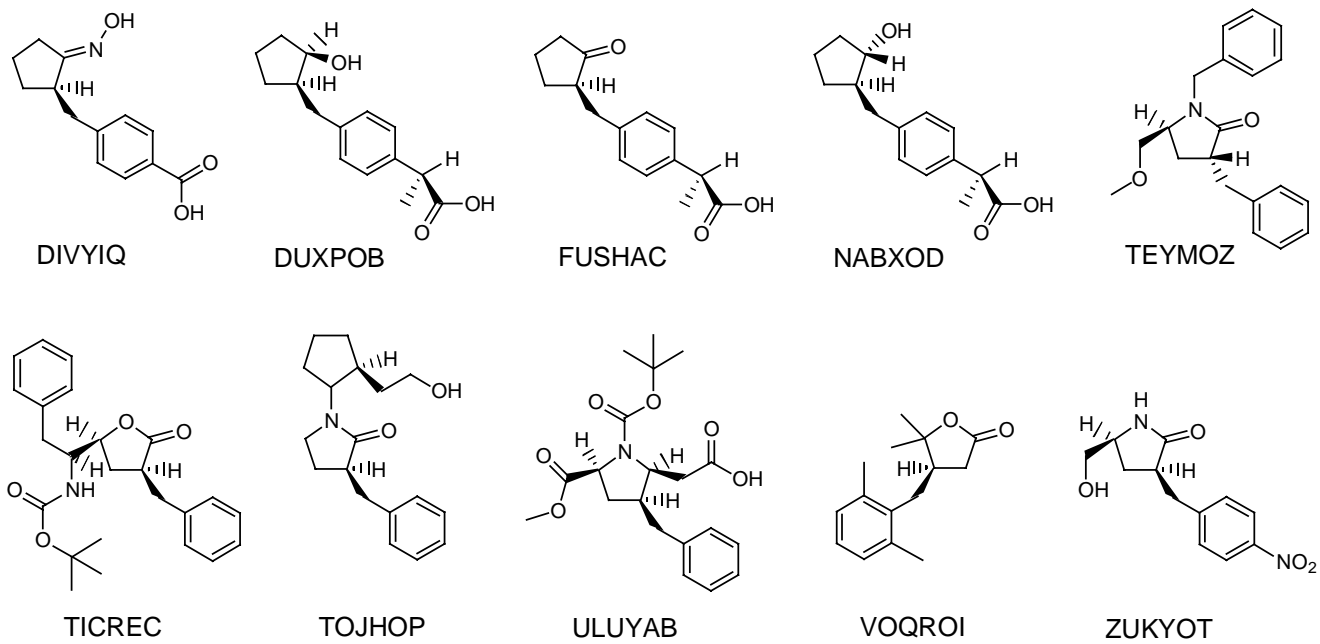


Figure C. The ten most similar structures to MDBL as retrieved from the CSD database, with their CSD codes.

Table C. Structural data for the ten MDBL-like compounds<sup>a,b</sup> from Figure C

Structure	$N_{CT}$	$N_{CN}$	$N_{CC}$	$N_{HB}$	$N_{HBC}$	$N_{in}$	$N_A$	$N_H$	$w_H/\%$	$M_r$	$v/\text{\AA}^3$	$\rho/\text{g cm}^{-3}$	$V/\text{\AA}^3$	$S/\text{\AA}^2$	$C_k$	$\mu/\text{cm}^{-1}$	$\Delta\theta$
DIVYIQ	67	12	2	5	4	28	35	17	6.9	247.29	322.1(3)	1.2750(10)	291.47	313.41	0.905	0.90	9.22°
DUXPOB-1	47	7	4	3	3	28	38	20	8.1	248.31	340.4(1)	1.2114(2)	300.99	319.45	0.884	0.83	1.62°
DUXPOB-2	59	9	4	3	3	32	38	20	8.1	248.31	340.4(1)	1.2114(2)	300.99	319.45	0.884	0.83	10.92°
FUSHAC	47	10	2	1	1	32	36	18	7.4	246.29	335.0(0.4)	1.2208(1)	293.44	308.22	0.876	0.84	4.58°
NABXOD*	61	12	4	3	2	26	38	20	8.1	248.31	332.4(1)	1.2405(3)	301.27	320.90	0.906	0.85	6.33°
TEYMOZ	38	14	0	3	0	44	46	23	7.5	309.39	431.6(2)	1.1900(6)	375.49	387.62	0.870	0.76	11.70°
TICREC	50	14	0	5	1	70	58	29	7.4	395.48	568.0(1)	1.1562(2)	470.96	481.13	0.829	0.78	14.21°
TOJHOP	56	14	0	0	0	42	42	22	8.1	274.36	378.2(3)	1.2048(8)	330.52	346.15	0.874	0.80	0.61°
ULUYAB	43	10	2	3	1	82	54	27	7.2	377.43	510.7(3)	1.2273(6)	439.07	458.91	0.860	0.90	6.62°
VOQROI	32	10	0	1	0	60	37	20	8.7	232.31	341.5(1)	1.1297(2)	286.32	293.31	0.838	0.73	0.62°
ZUKYOT-1	25	9	0	2	0	28	30	12	4.9	248.24	336.2(3)	1.2263(8)	275.88	286.14	0.821	0.94	5.55°
ZUKYOT-2	27	9	0	1	0	20	30	12	4.9	248.24	336.2(3)	1.2263(8)	275.88	286.14	0.821	0.94	3.55°

<sup>a</sup>Crystal structure is represented by its CSD code, as used in Table 2 and Figure C. Data multiplication for mirror-related structures was avoided, in contrary to Table 1.

<sup>b</sup>Structural descriptors are defined in the list below this table. Experimental errors are given for  $v$  and  $\rho$  and are related to the last digit.

\*Originally missing in the crystal structure, hydrogen atoms were modeled by using programs Mercury and Titan, sometimes aided by manual adjustment of the shortest hydrogen-hydrogen acceptor distances.

## Description of structural descriptors in Table C

1) Obtained from analysis of the crystal structures by the PLATON program:

$N_{CT}$  - total number of short intermolecular contacts that were counted using these criteria for general atoms I, J and X:

Distance  $d(I-J) < R(I) + R(J) + T$ , with tolerance  $T = 0.2 \text{ \AA}$

Angle  $\angle(X - I...J) > 100^\circ$

Contact (van der Waals) radii  $R$ :

Element	C	H	N	O
$R / \text{\AA}$	1.70	1.20	1.55	1.52

$N_{CN}$  - coordination number obtained from the above analysis

$N_{CC}$  - number of close non-hydrogen bond contacts from the above analysis

$N_{HB}$  - number of potential hydrogen bonds (classic and non-classic) satisfying following geometrical criteria for general donor (D), acceptor (A) and hydrogen (H) atoms:

Distance  $d(D...A) < R(D) + R(A) + 0.50 \text{ \AA}$

Distance  $d(H...A) < R(H) + R(A) - 0.12 \text{ \AA}$

Angle  $\angle(D-H...A) > 100^\circ$

By classic hydrogen bonds are understood bonds of the general type N,O-H...N,O.

Non-classic hydrogen bonds are weaker and of general type C-H...N,O.

$N_{HBC}$  - number of classic hydrogen bonds from the above hydrogen bond analysis

$N_{In}$  - number of short intramolecular contacts that were counted using the same criteria as the shortest intramolecular contacts.

$N_A$  - number of all atoms in a molecule

$N_H$  - number of hydrogen atoms in a molecule

$w_H$  - mass fraction of hydrogen atoms in a molecule

$M_r$  - relative molecular mass

$v$  - unit cell volume per molecule

$\rho$  - crystal density

$\mu$  - linear absorption coefficient of the crystal with respect to MoK $\alpha$  radiation

2) Obtained from PM3 optimized geometry and calculated by the Titan program:

$V$  - molecular volume

$S$  - molecular surface area

3) Combined descriptors:

$C_k$  -packing coefficient obtained as  $C_k = V/v$

$\Delta\Theta$  - absolute difference between crystallographic and computed (PM3) values for the rotation angle  $\Theta$  or  $\angle(C1'-C7'-C8'-H8')$

Two more descriptors were derived from these:

$N_S$  – number of the strongest intermolecular contacts defined as  $N_S = N_{HB} + N_{CC}$

$L(N_S)$  – linearized  $N_s$  as  $L(N_S) = (N_S - 5)^2$

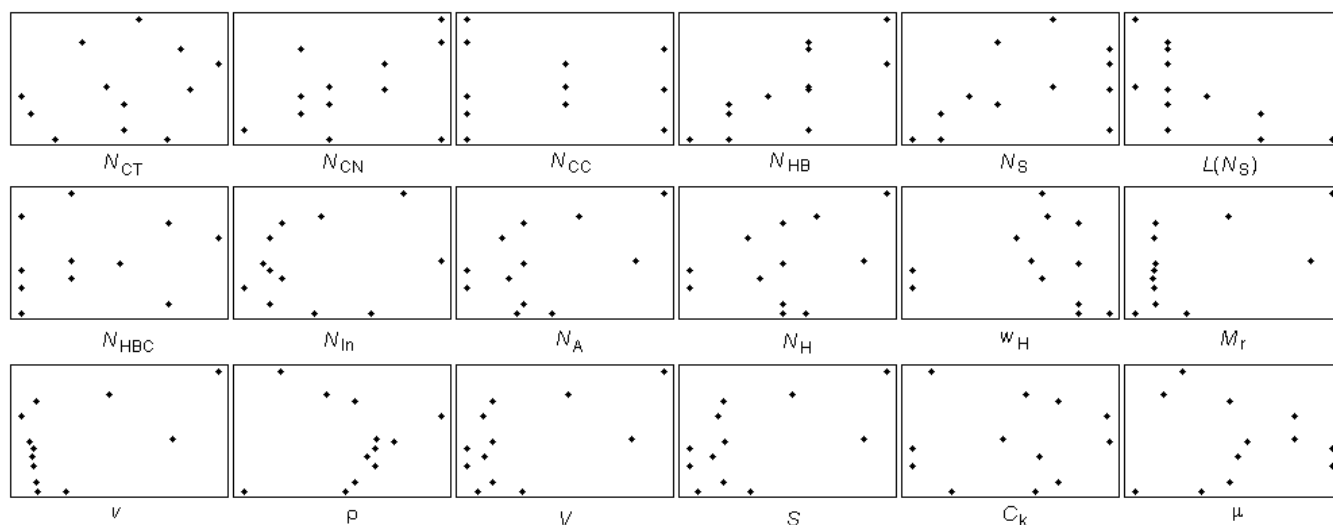


Figure D. Correlations of the structural descriptors with  $\Delta\Theta$  for the MDBL-like structures.

### Comments for Figure D

It is visible that some of the correlograms show regular trends, meaning that crystal packing and molecular structure determine the molecular conformation in the crystal and its computational approximation. Among good linear trends one can notice the  $\Delta\Theta$  increases with the elevation of the coordination number  $N_{CN}$  and the number of hydrogen bonds  $N_{HB}$ . The linearized  $N_S$  *i.e.*  $L(N_S)$  is also well-correlated with the rotation deviation.

There are also some interesting non-linear correlations of  $\Delta\Theta$  with the structural descriptors. The number of classic hydrogen bonds  $N_{HBC}$  shows that there are two groups of samples, each to be attributed to one of the two parallel parabolas in the correlograms, with maxima around 1.25. However, the correlograms for  $N_S$  exhibit its maximum at 5.

These two correlograms indicate that a classic hydrogen bond (N,O-H...N,O) and three non-classic hydrogen bonds (C-H...N,O,C) are responsible for the largest conformational change of an MDB-like molecule in crystal, because of what semi-empirical PM3 is generally unable to reproduce a close conformation. Smaller number of such hydrogen bonds can mean a conformation closer to that of the free state, while elevated number of the hydrogen bonds may cause some isotropically distributed intermolecular forces that mainly cancel each other.

The third type of illustrative correlograms is where two groups of samples show opposite trends and meet at extremely low or high values of the descriptor in question. Such descriptors are the number of the shortest intramolecular contacts  $N_{in}$ , the number of all atoms  $N_A$ , crystal density  $\rho$ , molecular volume  $V$  and surface area  $S$  and the linear absorption coefficient  $\mu$ . Other correlograms seem to consist of two or more sample groups, sometimes arranged as parallel lines, but due to the small number of samples it is impossible to evaluate the statistical significance of such trends. It is certain that symmetrically different molecules in the same asymmetric unit have distinct crystallographic environment, and consequently, their numbers of various contact types are not equal: see Table C for samples DUXPOB-1 and DUXPOB-2, and also for ZUKYOT-1 and ZUKYOT-2.

Numerical values associated to the correlograms in Figure C are in Table D.

Table D. Correlation coefficients between the descriptors from Figure D and the rotation deviation  $\Delta\Theta$

Descriptor	Correlation coefficient
$N_{CT}$	0.290
$N_{CN}$	<b>0.431</b>
$N_{CC}$	0.036
$N_{HB}$	<b>0.776</b>
$N_S$	0.480
$L(N_S)$	<b>-0.687</b>
$N_{HBC}$	0.260
$N_{in}$	0.217
$N_A$	0.500
$N_H$	0.404
$w_H$	-0.061
$M_r$	0.549
$v$	0.531
$\rho$	0.039
$V$	0.555
$S$	0.562
$C_k$	0.063
$\mu$	-0.062



## PLS analysis for data from Table C

Three descriptors with bold correlation coefficients in Table D were selected for PLS (Partial Least Squares regression) analysis to predict the deviations  $\Delta\Theta$ . The following results were obtained:

1 PC, 57.7% total variance, SEV = 2.78°, SEC = 2.44°, Q = 0.760, R = 0.853, maximal error 4.29°.

Regression vector:  $N_{\text{CH}}$ : 0.249       $N_{\text{HB}}$ : 0.449       $L(N_s)$ : -0.397

The predictions are illustrated in Figure E.

**Interpretation:** Molecular size and the number of various functional groups, especially hydrogen bonding and polar groups result in the increase of the coordination number and hydrogen bonds, which, with a few shortest non-hydrogen bonding contacts cause significant changes in molecular conformation. Such a conformation cannot be reproduced by PM3 and presumably with other semi-empirical methods, and it is supposed that higher computational levels (*ab initio*, DFT) would not be much more successful in this also. However, these computational procedures could be powerful if applied to molecular complexes as parts of the coordination sphere around the referent molecule. This PLS prediction is sufficiently quantitative to be used in deciding what computational strategy besides PM3 optimization of an isolated MDBL-like molecule is the most suitable.

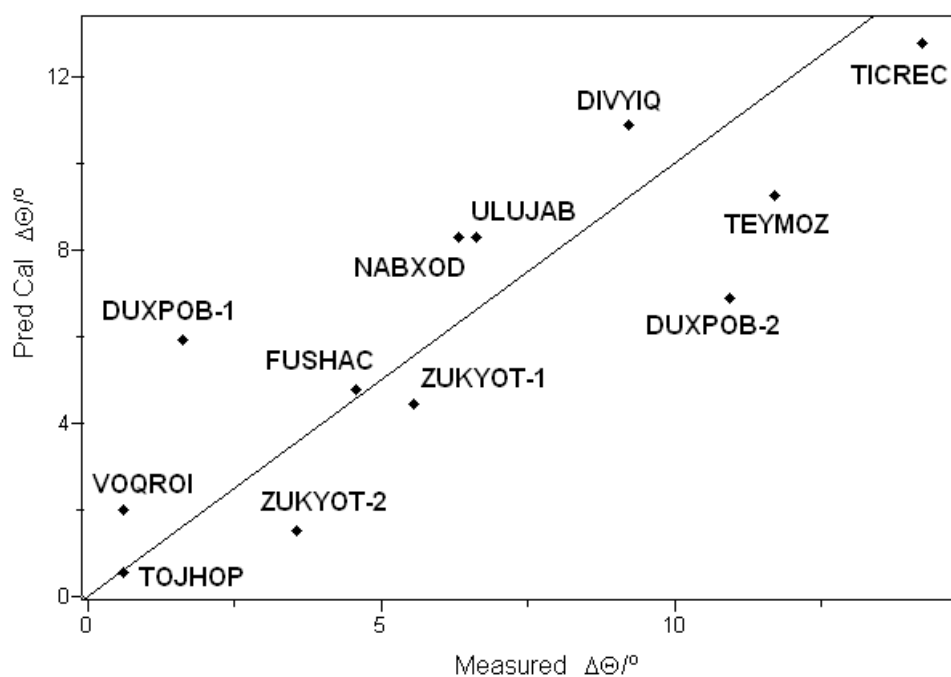


Figure E. PLS prediction of the PM3 rotation deviation  $\Delta\Theta$ .

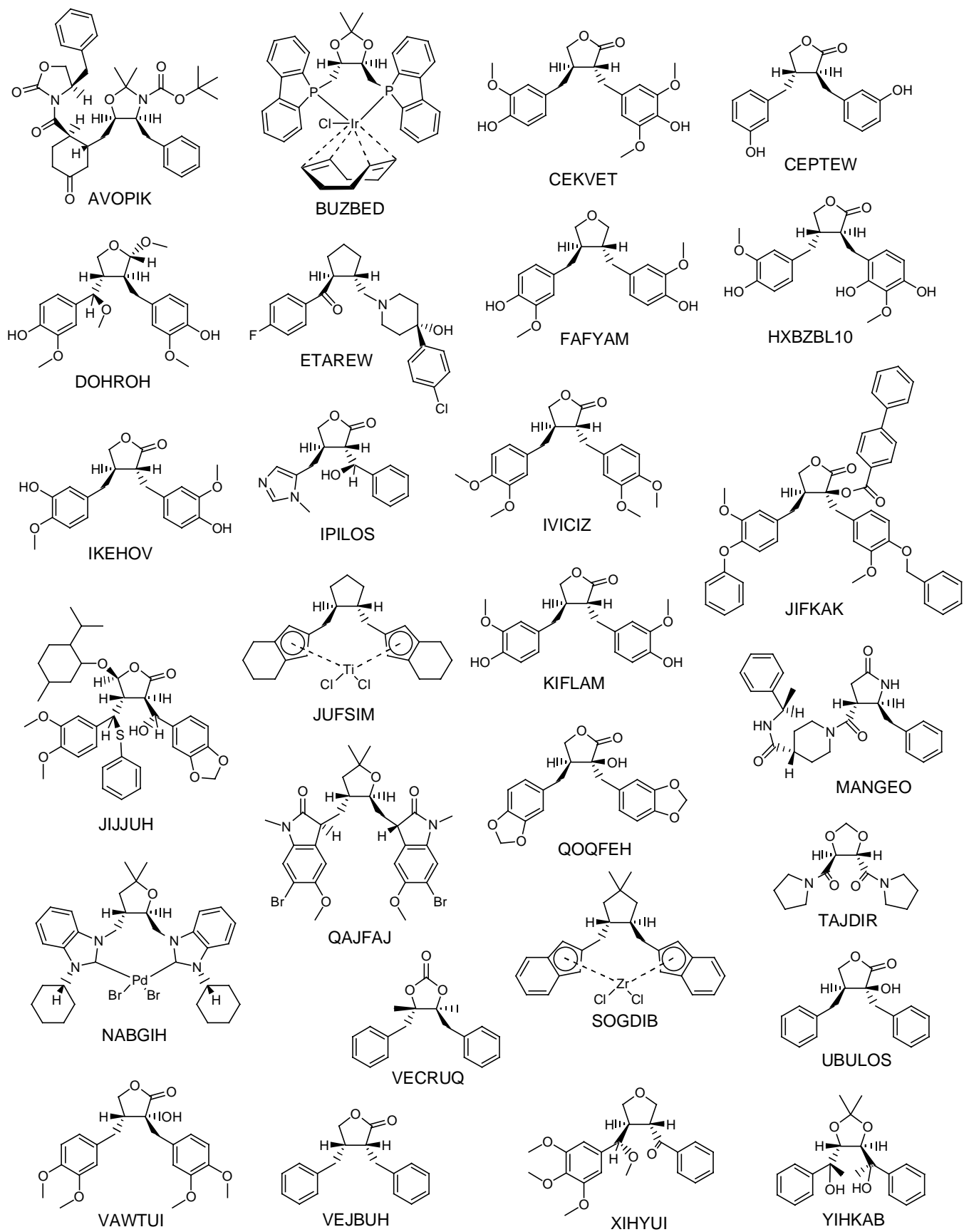


Figure F. The 27 HK-like structures retrieved from the CSD database with their CSD codes.

Table E. Structural data for the 27 HK-like compounds<sup>a,b</sup> from Figure F

Structure	$N_{CT}$	$N_{In}$	$N_{CN}$	$N_{CC}$	$N_{HBG}$	$N_{HBO}$	$N_{HBI}$	$N_{CT}$	$N_{CO}$	$N_{Cl}$	$N_A$	$N_H$	$w_H/\%$	$M_r$	$v/\text{\AA}^3$	$\rho/\text{g cm}^{-3}$	$V/\text{\AA}^3$	$S/\text{\AA}^2$	$C_k$	$\mu/\text{cm}^{-1}$	$\Delta\Theta$
AVOPIK-1	85	159	15	0	10	3	7	0	0	0	88	44	7.3	604.72	830.6(0.4)	1.2090(1)	694.31	664.65	0.836	0.84	26.15
AVOPIK-2	87	166	14	0	10	3	7	0	0	0	88	44	7.3	604.72	830.6(0.4)	1.2090(1)	694.31	664.65	0.836	0.84	37.97
BUZBED*	86	100	12	0	3	3	0	1	1	0	85	40	4.9	790.00	870.8(3)	1.5834(5)	723.80	665.90	0.831	40.35	28.04
CEKVET	104	92	12	0	7	4	3	4	2	2	52	24	6.2	388.40	469.0(1)	1.3752(3)	430.99	432.22	0.919	1.03	13.98
CEPTEW	63	32	13	4	4	4	0	2	2	0	40	18	6.1	298.32	384.3(1)	1.2890(3)	339.19	346.26	0.883	0.91	30.92
DOHROH	94	76	16	2	4	4	0	3	3	0	57	28	7.0	404.44	518.8(2)	1.2946(5)	458.18	461.66	0.883	0.96	4.87
ETAREW	36	66	10	1	3	2	1	2	2	0	56	27	6.5	415.92	535.7(3)	1.2892(6)	462.44	465.11	0.863	2.07	30.58
FAFYAM	16	50	6	4	3	2	1	2	1	1	49	24	7.0	344.39	445.0(0.4)	1.2851(11)	395.21	395.07	0.888	0.92	30.83
HXBZBL10*	85	72	12	2	8	5	3	5	3	2	49	22	5.9	374.38	460.0(3)	1.3516(7)	410.60	414.27	0.893	1.02	15.18
IKEHOV	92	60	13	0	4	2	2	3	1	2	48	22	6.2	358.38	434.0(1)	1.3714(2)	401.74	410.49	0.926	1.01	22.95
IPILOS*	68	44	14	2	3	3	0	1	1	0	39	18	6.3	286.32	366.6(5)	1.2969(18)	326.44	333.95	0.890	0.91	16.84
IVICIZ	108	74	14	0	5	4	1	0	0	0	54	26	6.8	386.43	478.32(1)	1.3415	443.59	460.03	0.927	0.97	8.23
JIFKAK	98	132	14	0	4	2	2	0	0	0	107	42	5.8	734.81	970.9(3)	1.2568(4)	826.55	808.30	0.851	0.85	22.02
JIJUH	72	127	14	0	7	4	3	0	0	0	89	43	6.7	647.78	857.5(6)	1.2545(8)	726.79	705.08	0.847	1.45	22.00
JUFSIM	40	148	10	0	3	0	3	0	0	0	60	32	7.1	451.28	554.3(2)	1.3519(5)	466.74	403.93	0.842	6.36	32.07
KIFLAM	35	45	10	4	2	1	1	2	1	1	49	24	7.0	344.39	446.9(1)	1.2797(39)	396.88	401.87	0.888	0.91	79.18
MANGEO	99	56	14	0	8	4	4	2	2	0	61	29	6.7	435.51	564.9(3)	1.2803(7)	502.89	517.36	0.890	0.88	66.83
NABGIH	52	126	15	0	1	1	0	0	0	0	99	42	5.7	885.06	955.65(6)	1.5739(1)	813.05	789.41	0.851	26.15	15.17
QAJFAJ	88	116	8	0	6	4	2	0	0	0	77	38	5.8	662.43	750.8(2)	1.4651(4)	639.00	598.24	0.851	27.37	41.66
QOQFEH	82	42	14	0	6	6	0	1	1	0	45	18	4.9	370.34	426.9(0.5)	1.4407(2)	387.87	386.10	0.909	1.10	12.41
SOGDIB	40	120	12	0	2	2	0	0	0	0	54	24	4.7	518.56	566.5(2)	1.5201(5)	484.21	460.22	0.855	7.40	8.00
TAJDIR	32	52	12	0	2	0	2	0	0	0	45	24	8.2	296.36	386.4(1)	1.2735(2)	347.48	364.24	0.899	0.92	35.41

UBULOS	38	46	10	0	1	1	0	1	1	0	39	18	6.4	282.32	373.3(5)	1.2560(17)	328.20	336.15	0.879	0.85	45.72
VAWTUI	108	88	13	0	5	5	0	2	2	0	53	26	6.5	402.43	507.9(4)	1.3158(10)	451.53	455.09	0.899	0.98	4.53
VECRUQ	19	33	10	0	0	0	0	0	0	0	42	20	6.8	296.35	402.5(2)	1.2225(7)	346.46	349.03	0.861	0.82	19.25
VEJBUH	28	42	12	0	1	1	0	0	0	0	38	18	6.8	266.32	366.0(3)	1.2082(9)	316.72	325.18	0.865	0.77	25.54
XIHYUI	66	73	12	0	5	3	2	0	0	0	54	26	6.8	386.43	495.93(4)	1.2939(1)	442.04	445.60	0.891	0.94	28.77
YIHKAB	50	82	10	2	4	1	3	2	1	1	51	26	7.7	342.42	480.46	1.1835	411.53	415.28	0.857	0.81	3.36

<sup>a</sup>Crystal structure is represented by its CSD code, as used in Table 3. Data multiplication for mirror-related structures was avoided, in contrary to Table 3.

<sup>b</sup>Structural descriptors are defined in the list below this table. Experimental errors are given for  $\nu$  and  $\rho$  and are related to the last digit.

\*Originally missing in the crystal structure, hydrogen atoms were modeled by using programs Mercury and Titan, sometimes aided by manual adjustment of the shortest hydrogen-hydrogen acceptor distances.

## Description of structural descriptors in Table E

Most descriptors are the same as defined in Table C, and were obtained in the same procedure.

The only new descriptors are hydrogen bond descriptors, obtained by analysis of crystal structures by means of PLATON as described for Table C.

$N_{\text{HBT}}$  - total number of potential hydrogen bonds (classic and non-classic), defined as in Table C

$N_{\text{HBO}}$  – number of potential intermolecular hydrogen bonds (classic and non-classic)

$N_{\text{HBI}}$  – number of potential intramolecular hydrogen bonds (classic and non-classic)

$N_{\text{CT}}$  – total number of classic hydrogen bonds from the above hydrogen bond analysis

$N_{\text{CO}}$  - number of classic intermolecular hydrogen bonds from the above hydrogen bond analysis

$N_{\text{CI}}$  - number of classic intramolecular hydrogen bonds from the above hydrogen bond analysis

A new descriptor was defined as  $L(N_{\text{HBT}}) = (N_{\text{HBT}} - 4)^2$ .

The rotation angle deviation  $\Delta\Theta$  was defined differently than before. Instead of including only one angle deviation (crystallographic – PM3) as for MDBL-like structures,  $\Delta\Theta$  is defined as the sum of these deviations for the two angles  $\Theta_1$  (C1'-C7'-C8'-H8') and  $\Theta_2$  (C1-C7-C8-H8). This is because of approximately mirror-related structures in most cases and the opposite correlogram trends observed for the individual angles.

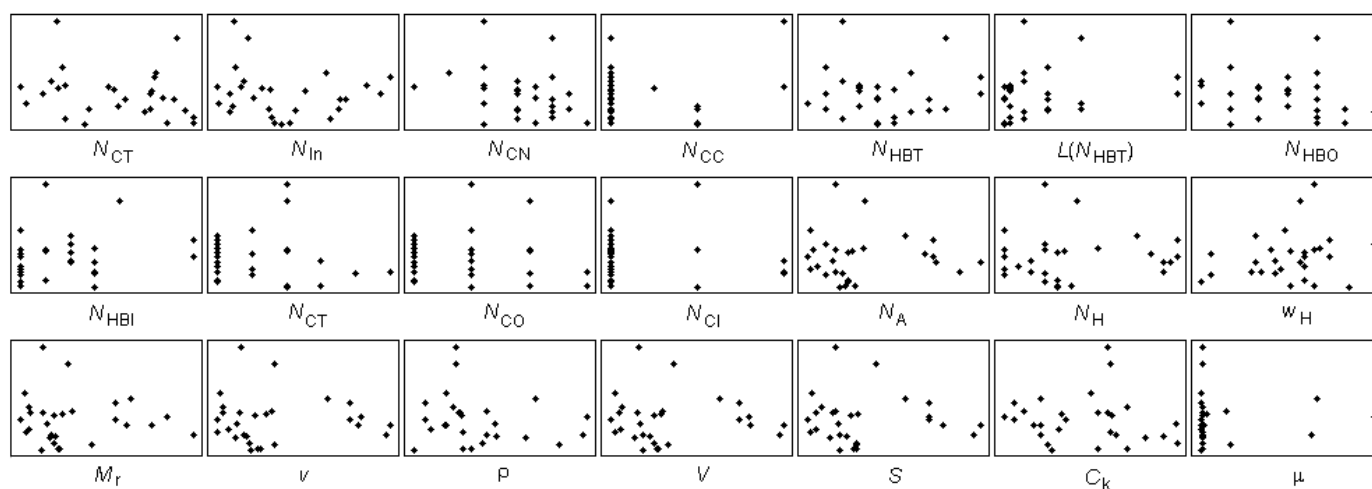


Figure G. Correlations of the structural descriptors with  $\Delta\Theta$  for the HK-like structures

## Comments for Figure G

Most correlograms in this figure do not show so clear trends like those in Figure D. This is due to the structure of the common skeleton of HK-like structures which is more complicated and has more conformational degrees of freedom than the skeleton of MDBL-like structures, and also because of large variations of structural subtypes among HK-like structures (what includes compounds with metals and large substituents). There are at least two groups of structures, obviously separated in the correlograms where  $\Delta\Theta$  is linearly correlated to a structural descriptor for each group:  $N_A$ ,  $N_H$ ,  $M_T$ ,  $\nu$ ,  $\rho$ ,  $V$  and  $S$ . All these descriptors illustrate that larger the molecule and though more dense the crystal, more similar is its conformation in crystal to that in the free state. It is like smaller molecules have more free space or are less tightly bound to each other in crystal so they can undergo more pronounced vibrations and disorders. Similar information may be obtained if one looks correlograms where the samples are rather spread but some linear tendency is retained: coordination number  $N_{CN}$  and the number of classic intermolecular hydrogen bonds  $N_{CO}$ .

There are some correlograms showing non-linear dependence of  $\Delta\Theta$  on a structural descriptor. While a parabola with minimum at low  $\Delta\Theta$  is not so clearly defined in the  $\Delta\Theta$  -  $N_{CC}$  plot, the analogous curve in the  $\Delta\Theta$  -  $N_{HBT}$  plot is much better defined. It is obvious that PM3 conformations for molecules with the minimum value  $N_{HBT} = 4$  will best reproduce the crystallographic conformations. However, this plot exhibits the samples in at least three groups. When this descriptor is linearized into  $L(N_{HBT}) = (N_{HBT} - 4)^2$ , the separation between three groups is more obvious. Correlation coefficients related to Figure G are in Table F below. Low correlations do not permit to build a regression model for prediction of  $\Delta\Theta$  values.

Table F. Correlation coefficients between the descriptors from Figure G and the rotation deviation  $\Delta\Theta$

Descriptor	Correlation coefficient
$N_{CT}$	-0.229
$N_{In}$	-0.140
$N_{CN}$	-0.304
$N_{CC}$	0.201
$N_{HBT}$	0.013
$L(N_{HBT})$	0.196
$N_{HBO}$	-0.242
$N_{HBI}$	0.219
$N_{CT}$	-0.073
$N_{CO}$	-0.068
$N_{CI}$	-0.052
$N_A$	0.006
$N_H$	0.062
$w_H$	0.212
$M_r$	-0.043
$\nu$	-0.007
$\rho$	-0.174
$V$	-0.015
$S$	-0.018
$C_k$	-0.091
$\mu$	0.028

Table G. The rotation angle deviation  $\Delta\Theta$  for a MDBL-like structure (CSD REFCODE: DIVJIQ<sup>a</sup>) obtained from different computational approaches<sup>b</sup>

Method <sup>c</sup>	Structure <sup>d</sup>	Target $\Delta\Theta^e$	$\Delta\Theta$ for other molecules <sup>e</sup>
MMFF94	single molecule	10.78°	-
Sybyl	single molecule	12.97°	-
(MD/MM2)/MMFF94 <sup>#</sup>	single molecule	14.55°	-
(MD/MM2)/Sybyl <sup>#</sup>	single molecule	15.90°	-
MNDO	single molecule	10.56°	-
AM1	single molecule	11.84°	-
PM3	single molecule	9.22°	-
HF: STO-3G	single molecule	15.65°	-
B3LYP: STO-3G	single molecule	9.12°	-
HF: 3-21G*	single molecule	12.04°	-
B3LYP: 3-21G*	single molecule	10.98°	-
HF: 6-31G*	single molecule	12.65°	-
B3LYP: 6-31G*	single molecule	12.90°	-
HF: 6-31G**	single molecule	12.68°	-
B3LYP: 6-31G**	single molecule	11.06°	-
<b>(MD/MM2)/MMFF94<sup>#</sup></b>	<b>HB dimer</b>	<b>4.71°</b>	<b>11.35°</b>
(MD/MM2)/MMFF94 <sup>#</sup>	HB dimer	13.91°	4.11°
MMFF94	HB dimer A	8.61°	10.56°
Sybyl	HB dimer A	17.69°	14.29°
<b>(MD/MM2)/MMFF94</b>	<b>HB dimer A</b>	<b>1.04°</b>	<b>16.71°</b>
(MD/MM2)/Sybyl	HB dimer A	20.34°	16.55°
MNDO	HB dimer A	12.69°	12.12°
AM1	HB dimer A	8.44°	12.08°
<b>PM3</b>	<b>HB dimer A</b>	<b>3.65°</b>	<b>8.56°</b>
B3LYP: STO-3G	HB dimer A	9.54°	7.03°
MMFF94	HB dimer B	15.00°	8.41°
Sybyl	HB dimer B	13.53°	13.26°
(MD/MM2)/MMFF94	HB dimer B	12.24°	14.71°
(MD/MM2)/Sybyl	HB dimer B	16.83°	16.00°
MNDO	HB dimer B	12.36°	11.18°
AM1	HB dimer B	8.50°	12.25°
PM3	HB dimer B	8.90°	8.72°
B3LYP: STO-3G	HB dimer B	9.54°	7.03°
(MD/MM2)/MMFF94 <sup>#</sup>	HB trimer	15.52°	12.72°, 17.66°
(MD/MM2)/MMFF94 <sup>#</sup>	HB trimer	18.79°	12.93°, 14.88°
MMFF94	HB hexamer	14.56°	8.44°, 9.75°, 12.87°, 14.12°, 18.07°
Sybyl	HB hexamer	13.15°	13.04°, 13.88°, 14.81°, 16.02°, 16.48°
<b>(MD/MM2)/MMFF94</b>	<b>HB hexamer</b>	<b>3.04°</b>	<b>0.44°, 10.88°, 20.55°, 27.84°, 37.91°</b>
(MD/MM2)/Sybyl	HB hexamer	6.88°	9.41°, 11.52°, 18.76°, 21.74°, 22.90°
MNDO	HB hexamer	12.16°	11.02°, 11.75°, 12.06°, 12.07°, 12.08°
AM1	HB hexamer	11.56°	9.10°, 12.02°, 12.27°, 12.42°, 12.48°
PM3	HB hexamer	12.69°	2.62°, 6.43°, 9.02°, 10.61°, 12.39°
MMFF94	Coord. Sphere	5.58°	1.02°, 6.47°, 14.12°, 13.08°, 15.68°, 18.52°, 19.48°, 25.58°, 28.80°, 31.34°, 42.84°, 43.71°
Sybyl	Coord. Sphere	12.10°	3.11°, 7.29°, 8.27°, 9.00°, 9.50°, 9.52°, 15.02°, 17.87°, 21.00°, 22.13°, 22.30°, 26.11°
<b>(MD/MM2)/MMFF94</b>	<b>Coord. Sphere</b>	<b>4.30°</b>	<b>3.98°, 10.71°, 11.16°, 12.79°, 14.40°, 15.23°, 15.86°, 18.76°, 19.32°, 21.75°, 24.30°, 36.99°</b>
(MD/MM2)/Sybyl	Coord. Sphere	11.99°	0.57°, 1.08°, 6.19°, 7.10°, 7.39°, 9.43°, 11.76°, 12.45°, 13.66°, 16.98°, 23.75°, 36.18°



<sup>a</sup>This compound is one of the ten MDBL-like structures retrieved from the Cambridge Structural Database. It is (*E*)-4-(2-hydroxyiminocyclopentylmethyl)-phenylacetic acid C<sub>14</sub>H<sub>17</sub>NO<sub>3</sub>, with crystal structure determined by X-ray CuK $\alpha$  radiation at 298 K and factor  $R=5.97\%$  [T. Hatu, S. Sato, C. Tamura. (*E*)-4-(2-hydroxyiminocyclopentylmethyl)-phenylacetic acid. *Acta Cryst.* **C42** (1986) 452-454.] It is a colorless crystal that crystallizes in space group  $P2_1/c$  and has four molecules in the unit cell (two enantiomeric pairs). <sup>b</sup>Computational procedures were carried out using programs Titan [Titan v. 1.0.8., Wavenfunction Inc. & Schrödinger Inc., 2001] and Chem3D [CS Chem3D Ultra, CambridgeSoft.Com, Cambridge MA, 2000]. The computational approaches included: molecular mechanics force fields MMFF94 and Sybyl in the Titan software, molecular dynamics (MD) and molecular mechanics force field MM2 in the Chem3D software, semi-empirical (MNDO, AM1, PM3), *ab initio* (HF) and DFT (B3LYP) methods in the Titan software. Combined approaches as (MD/MM2)/MMFF94 stand for MD (default parameters) followed with MM2 geometry optimization and then with MMFF94 treatment. <sup>c</sup>All the methods used experimental structure for the initial geometry, except for methods marked with # that were used first to draw the structures and then treat them computationally. <sup>d</sup>Besides single molecule, molecular aggregates were also considered. Hydrogen bonded dimers A and B differ in the neighboring molecule in interaction with the referent molecule (the molecule that was treated with purpose to minimize  $\Delta\Theta$ ). Hydrogen bonding hexamer includes five molecules that are in interaction with the referent molecule mainly *via* hydrogen bonds. Coordination sphere includes the referent molecule and all twelve molecules that make the first coordination sphere around it. <sup>e</sup>The  $\Delta\Theta$  values for the referent molecule and other molecules from molecular aggregates. The bold methods are those representing the best results ( $\Delta\Theta < 5^\circ$ ) for the referent molecule.

## Comments for Table G

The  $\Delta\Theta$  results presented in Table E for the selected MDBL-like structure show some interesting trends that should be commented. The diverse calculations were performed in order to show the effect of crystal packing to the conformational parameter  $\Theta$  of this compound, as well as to point out the complexity of computational efforts to minimize  $\Delta\Theta$ . Although most of time *ab initio* and DFT are considered as the methods that in the best way reproduce experimental geometry, in case of conformational problems it can be rather different. Excellent reproducing of bond lengths and angles as well as of torsion angles defining closed structures (rings), in which *ab initio* and DFT methods are known to be superior to semi-empirical and empirical methods, is not considered here. What is of primary interest in this study is to reproduce  $\Theta$  values *i.e.* to minimize  $\Delta\Theta$  because the torsion angle  $\Theta$  defines the overall molecular conformation of the studied compound. Maybe higher levels of *ab initio* or DFT theory really could improve the results but such approaches would be time-expensive. Besides, there is no guarantee for better results anyway because two different phases are considered: a real crystal and the gas phase or very small molecular aggregation that loses the original structural features determined by the crystal symmetry.

The investigated compound's  $\Theta$  value is by some  $15^\circ$  smaller than the value of  $60^\circ$  for ideal *gauche* conformation. Most computationally obtained single molecules have angles rather close to this ideal value. However, crystal packing forces around the molecule, primarily hydrogen bonds, then crystal symmetry, the presence of identical and mirror-related molecules, and in some extent intramolecular interactions, define the experimental value of  $\Theta$ . It is very difficult or even impossible to simulate all these real conditions without some software specially designed for infinite systems, polymers and crystals. The energy minimum conformation of a single molecule is not necessary equal to the experimental conformation. Dimer, trimer, hexamer or even coordination sphere are not equal to the crystal environment, and their constitutive conformers will participate in intermolecular interactions that have lost the periodicity and symmetry from the crystal. It is obvious for dimers, trimers, hexamers and the complete coordination sphere that chemically equal molecules substantially differ in  $\Delta\Theta$  values, what does not happen in the crystal (unless there are more than one molecules in the asymmetric unit). It is impossible to predict which method will give the smallest  $\Delta\Theta$  value for the referent molecule and not for other molecules. Some general conclusion can be made that molecular aggregations may yield better results at some molecular mechanics (in this example: MMFF94) or even semi-empirical (in this example: PM3) methods. Probably some other example would not indicate exactly the same particular methods as the best ones. Therefore, a simple semi-empirical treatment (PM3) applied to experimental structures of selected MDBL-like and HK-like compounds seems to be a reasonable method that does not produce heterogeneous results in this work. The effect of crystal packing and symmetry on molecular conformation has been finally proven by applying the systematic computational procedures as showed in Table G.

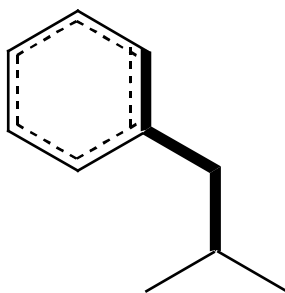


Figure H. Fragment defining the torsion angle CH-C-CH<sub>2</sub>-C (along the bold lines) that describes the relative position of the phenyl group with respect to a general alkyl group. This fragment was used in the searches of Cambridge Structural Database (CSD), without any filtering.