An Improved experimental databank of transferable multipolar atom models - ELMAM2. Construction details and applications

Sławomir Domagała, Bertrand Fournier, Dorothee Liebschner, Benoît Guillot[a] & Christian Jelsch[*]

Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2), CNRS, UMR 7036, Institut Jean Barriol, Faculté des Sciences et Techniques, Nancy Université, BP 70239, 54506 Vandoeuvre-lès-Nancy CEDEX, France

E-mail: christian.jelsch@crm2.uhp-nancy.fr

SUPPORTING INFORMATION

Supplementary Tables 1S-6S

Supplementary Figures 1S

Supplementary Tables

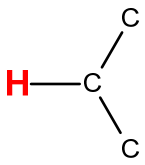
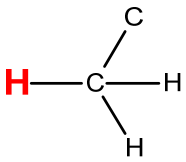
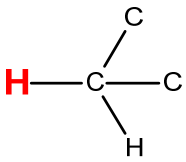
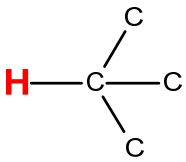
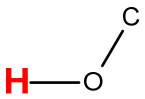
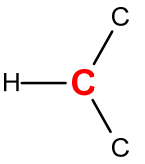
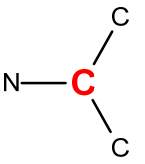
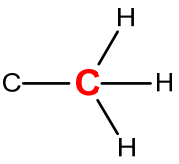
Table 1S. List of crystal structures used for building the ELMAM2 database.

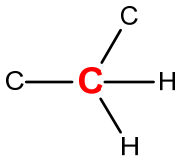
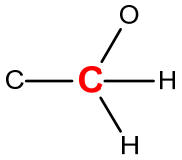
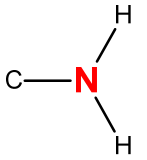
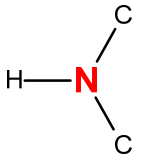
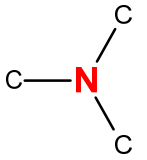
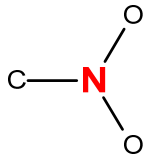
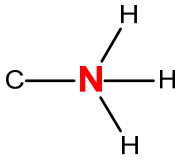
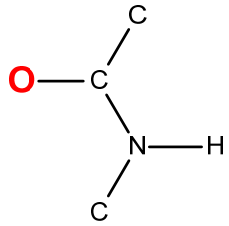
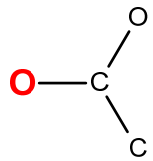
Nr	Abbreviation	Name, formula	Data type, temp	Reference
1	ala	L-Alanine, $C_3H_7NO_2$	X-Ray, Mo, 23K	Destro *et al.*, 1988
2	enkephalin	Leu-Enkephalin Trihydrate, $C_{28}H_{37}N_5O_7$ $3H_2O$	X-Ray, Mo, 100K	Wiest *et al.*, 1994
3	Argphos	L-Arginine Phosphate Monohydrate, $C_6H_{15}N_4O_2^+$ $H_2PO_4^-$ H_2O	X-Ray, Mo, 130K	Espinosa *et al.*, 1996
4	oxalic	Oxalic acid, $C_2H_2O_4$	X-Ray, Mo, 100K	Martin *et al.*, 1998
5	his	DL-Histidine, $C_6H_9N_3O_2$	X-Ray, Mo, 110K	Coppens *et al.*, 1999
6	actyr	N-acetyl-L-tyrosine ethyl ester monohydrate, $C_{13}H_{19}NO_5$	X-Ray, Mo, 110K	Dahaoui *et al.*, 1999
7	glythr	Glycyl-L-threonine dihydrate, $C_6H_{12}N_2O_4$ $2H_2O$	X-Ray, Mo, 110K	Benabicha *et al.*, 2000
8	GlyAsp	glycyl-aspartic acid dihydrate, $C_6H_{10}N_2O_5$ $2H_2O$	X-Ray, Mo, 123K	Pichon-Pesme *et al.*, 2000
9	TyrGlyGly	Tyrosyl-glycyl-glycine monohydrate, $C_{13}H_{17}N_3O_5$ H_2O	X-Ray, Mo, 123K	Pichon-Pesme *et al.*, 2000
10	panb	*p*-amino-*p'*-nitrobiphenyl, $C_{12}H_{10}N_2O_2$	X-Ray, Mo, 20K	Volkov *et al.*, 2000
11	ureaphos	urea-phosphoric acid, $CH_7N_2O_5P$	X-Ray, Mo, 100K	Rodrigues et al., 2001
12	NTO	5-nitro-2,4-dihydro-3H-1,2,4-triazol-3-one, □-NTO, $C_2H_2N_4O_3$	X-Ray, Ag, 100K	Zhurova *et al.*, 2001
13	cycloproala	cyclo-(D,L-Pro)$_2$-(L-Ala)$_4$ monohydrate, $C_{22}H_{34}N_6O_6$ H_2O	Synchrotron, 0.5583 Å, 100K	Dittrich *et al.*, 2002
14	Thiocoumarin 2	2H-chromene-2-thione, (2-thiocomarin) C_9H_6OS	X-Ray, Mo, 90K	Munshi *et al.*, 2002
15	HPPM	hydrogen-[(2,4-diaminopyrimidin-1-io)methyl]-phosphonate monohydrate, $C_5H_9N_4O_3P$ H_2O	X-Ray, Mo, 105K	Slouf *et al.*, 2002
16	BIGH	Biguanidinium Dinitramide, $C_2H_8O_4N_8$	X-Ray, Mo, 90K	Zhurova *et al.*, 2002
17	BIGH2	Biguanidinium Bis-Dinitramide, $C_2H_9O_8N_{11}$	X-Ray, Mo, 90K	Zhurova *et al.*, 2002
18	schiff1	dianil of 2-hydroxy-5-methyl-isophth-aldehyde, $C_{21}H_{18}O_1N_2$	X-Ray, Mo, 100K	Dominiak *et al.*, 2003
19	schiff2	3,5-dinitro-*N*-salicylidenoethylamine, $C_9H_9O_5N_3$	X-Ray, Mo, 100K	Dominiak *et al.*, 2003
20	schiff3	3-nitro-*N*-salicylidenocyclohexylamine, $C_{13}H_{16}O_3N_2$	X-Ray, Mo, 100K	Dominiak *et al.*, 2003
21	nad	NAD+, *β*-Nicotinamide Adenine Dinucleotide, $C_{21}H_{27}N_7O_{14}P_2$ $4H_2O$	X-Ray, Mo, 100K	Guillot et al., 2003

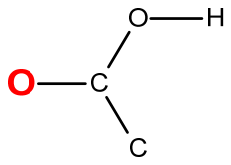
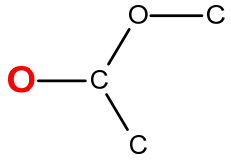
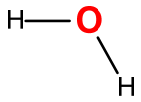
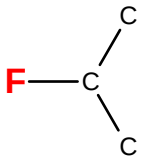
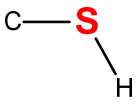
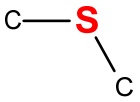
22	TFT	Tetrafluoroterephthalonitrile, $C_8N_2F_4$	X-Ray, Mo, 122.4K	Sørensen *et al.*, 2003
23	urea	Urea, CH_4N_2O	Synchrotron, 0.5996 Å, 123K	Birkedal *et al.*, 2004
24	busulfan	1,4-butanediol-dimethylsulfonate, $C_6H_{14}O_6S_2$	X-Ray, Mo, 100K	Ghermani *et al.*, 2004
25	barba	2,4,6,8-tetraphenylbarbaralane, $C_{33}H_{26}$	Synchrotron, 0.56 Å, 110K	Luger et al., 2004
26	xylitol	xylitol, $C_5H_{12}O_5$	X-Ray, Mo, 122.4 K	Madsen *et al.*, 2004
27	famotidineA	Famotidine (polymorph A), $C_8H_{15}N_7O_2S_3$	X-Ray, Mo, 100K	Overgaard *et al.*, 2004
28	famotidineB	Famotidine (polymorph B), $C_8H_{15}N_7O_2S_3$	X-Ray, Mo, 100K	Overgaard *et al.*, 2004
29	trp	L-tryptophan formic acid, $C_{11}H_{12}N_2O_2\ CH_2O_2$	Synchrotron, 0.54 Å, 100K	Scheins *et al.*, 2004
30	coumarin	2H-chromene-2-one, $C_9H_6O_2$	X-Ray, Mo, 90K	Munshi *et al.*, 2005
31	Thiocoumarin 1	1-thiocumarin, C_9H_6OS	X-Ray, Mo, 90K	Munshi *et al.*, 2005
32	pentaerythritol	Pentaerythritol, $C_5H_{12}O_4$	X-Ray, Mo, 15K	Zhurov *et al.*, 2005
33	accoumarinA	3-Acetylcoumarin, (polymorph A), $C_{11}H_8O_3$	X-Ray, Mo, 90K	Munshi *et al.*, 2006a
34	accoumarinB	3-Acetylcoumarin, (polymorph B), $C_{11}H_8O_3$	X-Ray, Mo, 90K	Munshi *et al.*, 2006a
35	cytosine_hyd	cytosine monohydrate, $C_4H_5N_3O\ H_2O$	X-Ray, Mo, 90K	Munshi *et al.*, 2006b
36	salicylic	salicylic acid, $C_7H_6O_3$	X-Ray, Mo, 90K	Munshi et al., 2006b
37	thiosemicarb	1-formyl-3-thiosemicarbazide, $C_2H_5N_3OS$	X-Ray, Mo, 90K	Munshi et al., 2006c
38	thiouracil	2-thiouracil, $C_4H_4N_2OS$	X-Ray, Mo, 90K	Munshi et al., 2006b
39	stilbene	Dimethylstilbene	X-Ray, 16K	Ogawa *et al.*, 2006
40	estradiol	17β-Estradiol Urea, $C_{18}H_{24}O_2\ CH_4N_2O$	X-Ray, Mo, 100K	Parrish *et al.*, 2006
41	estrone	3-hydroxy-1,3,5-(10)-estratrien-17-one, $C_{18}H_{22}O_2$	X-Ray, Mo, 20K	Zhurova *et al.*, 2006
42	diazapentalene	1,3,4-trinitro-7,8-diazapentalene, $C_6H_3N_5O_6$	X-Ray, Ag, 90K	Chen *et al.*, 2007
43	sarcosine	Sarcosine, $C_3H_7NO_2$	X-Ray, Mo, 100K	Dittrich *et al.*, 2007
44	alaproala	L-alanyl-L-prolyl-L-alanine hydrate, $C_{11}H_{19}N_3O_4\ H_2O$	Synchrotron, 100K	Kalinowski *et al.*, 2007
45	ornithine	L-ornithine hydrochloride, $C_5H_{13}N_2O_2\ Cl$	X-Ray, Mo, 100K	Dittrich *et al.*, 2007
46	thymidine	Thymidine, $C_{10}H_{14}N_2O_5$	X-Ray, Mo, 20K	Hübschle *et al.*, 2008
47	thiouronium	2-(indol-3-yl)-1,1,3,3-tetramethylthiouronium nitrate, $C_{13}H_{18}N_3S^+NO_3^-$	X-Ray, Mo, 110K	Lutz *et al.*, 2008
48	FOX-7	1,1-diamino-2,2-dinitroethylene, $C_2H_4N_4O_4$	Synchrotron, 0.503 Å, 100K	Meents *et al.*, 2008
49	RhPCl	[Rh(C7H8)(PPh3)Cl], $C_{25}H_{23}ClPRh$	X-Ray, Mo, 100K	Sparkes *et al.*, 2008
50	paracetamol	paracetamol, $C_8H_9NO_2$	X-Ray, Mo, 100K	Bouhmaida *et al.*, 2009

51	16Ni	[3,11-Bis(methoxycarbonyl)-1,5,9,13-tetraazacyclohexadeca-1,3,9,11-tetraenato-(2-)-κ^4N]nickel(II), $C_{16}H_{22}N_4O_4Ni$	X-Ray, Mo, 100K	Domagała *et al.*, 2009
52	fidarestat	Fidarestat, $C_{12}H_{10}N_3O_4F$	Synchrotron, 0.450 Å, 100K	Fournier *et al.*, 2009
53	coumcarb	coumarin-3-carboxylic acid, $C_{10}H_6O_4$	X-Ray, Mo, 100K	Howard *et al.*, 2009
54	transcin	trans-cinnamic acid, $C_9H_8O_2$	X-Ray, Mo, 100K	Howard *et al.*, 2009

Table 2S. Schematic representation of the selected, available atom-types in the ELMAM2.

Symbol	Chemical connectivity	Schematic representation
H101	Hc[cc]	
H103	Hc[chh]	
H104	Hc[cch]	
H105	Hc[ccc]	
H115	Ho[c]	
C301	C1.5c1.5c1h	
C307	C1.5n1.5c1.5c	
C401	C1c1h1h1h	

C402	C1c1c1h1h	
C410	C1o1c1h1h	
N301	Nchh	
N302	N1.5c1.5c1h	
N304	Nccc	
N305	Nooc	
N401	Nchhh	
O102	O1.5c[1.5n1c]	
O104	O1.5c[1.5o1c]	

O105	O2c[1o(1h)1c]	
O106	O2c[1o(1c)1c]	
O201T	Ohh	
F101T	Fc[cc]	
S201T	Sch	
S202	Scc	

The connectivity is shown without the bond orders. In total 68 atom-types are available; 16 H; 31C; 7N; 11O; 1F; 2S

Table 3S. Benchmarked crystal structures.

Compound	Structure details	Theoretical calculations	Database transfer details	Reference
Glycine	high resolution X-Ray, 23K	Single Point	AMBER (a mixture of C-terminal glycine, N-terminal glycine and bulk glycine used for CA, HA1 and HA2 point charges) ELMAM ELMAM2	Destro, R.; Roversi, P.; Barzaghi, M.; Marsh, R. E. *J. Phys. Chem. A*, **2000**, *104*, 1047-1054.
AlaGlyAla anhydrous	X-Ray, 293K	Single Point, X-H distances extended to average neutron values	AMBER ELMAM ELMAM2	Sreekanta Padiyar, G.; Parthasarathy Seshadri, T. *Acta Crystallogr., Sect. C*, **1996**, *52*, 1693-1695.
AlaGlyAla monohydrate	X-Ray, 20K	Single Point, X-H distances extended to average neutron values	AMBER ELMAM ELMAM2	Förster, D.; Messerschmidt, M.; Luger, P. *Acta Crystallogr., Sect. C*, **2005**, *61*, o420-o421.
AlaProAla monohydrate	high resolution, X-Ray, 100K	Single Point	AMBER ELMAM ELMAM2	Kalinowski, R.; Dittrich, B.; Hübschle, C. B.; Paulmann, C.; Luger, P. *Acta Crystallogr., Sect. B*, **2007**, *63*, 753-767.
AlaTyrAla ethanol	X-Ray, 20K	Single Point, X-H distances extended to average neutron values	AMBER ELMAM ELMAM2	Chęcińska, L.; Förster, D.; Morgenroth, W.; Luger, P. *Crystallogr., Sect. C*, **2006**, *62*, o454-o457.
benzene	Neutron, 138K	Single Point	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Bacon, G. E.; Curry, N. A.; Wilson, S. A. *Proc. R. Soc. London Ser. A*, 1964, 279, 98-110.
catechol	X-Ray, 100K	Full Optimization	ELMAM (tyrosine aromatic ring used to mimic atom-types)	Fronczek, F. R.; Kim, K. K.; Strongin, R. M.; *Priv. Comm.*, **2002**.

			ELMAM2	(CATCOL13)
resorcinol	Neutron, 298K	Single Point	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Bacon, G. E.; Jude R. J. *Z.Kristallogr., Kristallgeom., Kristallphys., Kristallchem.* **1973**, *138*, 19-40.
***p*-nitrophenol α-form**	high resolution X-Ray, 110K	Single Point	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Kulkarni, G. U.; Kumaradhas, P.; Rao, C. N. R. *Chem. Mater.* **1998**, *10*, 3498-3505.
***p*-nitrophenol β-form**	high resolution X-Ray, 110K	Single Point	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Kulkarni, G. U.; Kumaradhas, P.; Rao, C. N. R. *Chem. Mater.* **1998**, *10*, 3498-3505.
***p*-nitroaniline**	X-Ray, 123K	hydrogen atoms positions optimized	ELMAM2	Nieger, M. *Priv. Comm.* **2007**. (NANILI23)
***p*-nitrobenzoic acid**	X-Ray, 123K	hydrogen atoms positions optimized	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Groth, P. *Acta Chem. Scand. A* **1980**, *334*, 229-230.
2,5-dihydroxybenzoic acid	X-Ray, 130K	hydrogen atoms positions optimized	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Cohen, D. E.; Benedict, J. B.; Morlan, B.; Chiu, D. T.; Kahr, B. *Cryst. Growth Des.* **2007**, *7*, 492-495.
***p*-dinitrobenzene**	Neutron, 120K	Single Point	ELMAM (tyrosine aromatic ring used to mimic atom-types) ELMAM2	Tonogaki, M.; Kawata, T.; Ohba, S.; Iwata, Y.; Shibuya, I. *Acta Crystallogr., Sect. B* **1993**, *49*, 1031-1039.
quercetin monohydrate	X-Ray, 100K	Full Optimization	ELMAM2	Domagała, S., Munshi, P., Ahmed, M., Guillot, B. & Jelsch, C. (2011). *Acta Crystallogr., Sect. B* **2011**, *67*, 63-78.

Table 4S. Cross-validation statistics for the L-SerVal structure. The average values over the 20 working and test sets are given. All the *R*-factor values are in %.

Model	$wR^2(F)$	$wR^2(F)_{free}$	$S(F)$	$S(F)_{free}$
IAM_U_{iso}	4.15	5.20	2.55	2.92
IAM_U_{anis}	5.10	5.33	2.62	2.99
ELMAM_U_{iso}	2.47	2.95	1.52	1.65
ELMAM_U_{anis}	2.52	3.00	1.54	1.67
ELMAM2_U_{iso}	2.53	3.07	1.55	1.71
ELMAM2_U_{anis}	2.40	2.87	1.47	1.60

Electrostatic Potential quantities computed at the van der Waals surface for tri-peptides and aromatic molecules.

All the notations used here to describe the descriptors are the same as in the original papers (Murray & Politzer, 1998; Murray *et al.*, 2000).

The quantities: $V_{S,\max}$, $V_{S,\min}$, $\overline{V}_S^+$, $\overline{V}_S^-$, $\overline{V}_S$ and Π are given in kcal/mol; σ_+^2, σ_-^2 and σ_{tot}^2 are in (kcal/mol)^2; ν is dimensionless. V_S^+ and V_S^- designate the regions of positive and negative potential respectively. The following definitions were used to calculate surface quantities:

$$\overline{V}_S^+ = \frac{1}{p}\sum_{i=1}^{p} V_S^+(\mathbf{r}_i), \qquad \overline{V}_S^- = \frac{1}{n}\sum_{i=1}^{n} V_S^-(\mathbf{r}_i), \qquad \overline{V}_S = \frac{1}{p+n}\sum_{i=1}^{p+n} V_S(\mathbf{r}_i),$$

$$\Pi = \frac{1}{p+n}\sum_{i=1}^{p+n}\left|V_S(\mathbf{r}_i) - \overline{V}_S\right|, \qquad \sigma_+^2 = \frac{1}{p}\sum_{i=1}^{p}[V_S^+(\mathbf{r}_i) - \overline{V}_S^+]^2, \qquad \sigma_-^2 = \frac{1}{n}\sum_{i=1}^{n}[V_S^-(\mathbf{r}_i) - \overline{V}_S^-]^2,$$

$$\sigma_{\text{tot}}^2 = \sigma_+^2 + \sigma_-^2, \quad \nu = \frac{\sigma_+^2\sigma_-^2}{(\sigma_{\text{tot}}^2)^2}.$$

Table 5S. Electrostatic Potential quantities computed at the van der Waals surface for glycine and tri-peptides.

Surface quantity	AMBER	ELMAM	ELMAM2	THEO	AMBER	ELMAM	ELMAM2	THEO
glycine					**AGA**			
$V_{S,\max}$	125.1	102.8	86.3	91.4	153.2	119.8	103.0	118.5
$V_{S,\min}$	-122.3	-102.0	-77.9	-78.9	-159.2	-138.9	-105.8	-110.0
$\overline{V}_S^+$	59.7	52.8	44.2	45.0	61.3	49.7	40.4	49.8
$\overline{V}_S^-$	-63.5	-57.5	-44.7	-46.3	-63.2	-62.9	-49.3	-49.9
$\overline{V}_S$	1.9	2.1	4.7	5.1	-2.9	-6.6	-0.9	-0.6
Π	61.4	54.8	44.0	45.0	62.2	56.5	44.6	49.8
σ_+^2	1055.0	566.7	467.5	477.8	1360.4	810.7	510.8	697.1
σ_-^2	872.9	635.8	377.7	383.2	1611.9	1555.3	948.2	1032.7
σ_{tot}^2	1927.9	1202.5	845.2	861.0	2972.3	2366.0	1459.1	1729.8
ν	0.248	0.249	0.247	0.247	0.248	0.225	0.228	0.241
RMS	68.9	60.2	49.0	50.1	73.3	66.3	52.1	57.9
RMSD	20.7	12.9	2.8		20.4	15.4	8.9	
r	0.992	0.992	0.999		0.979	0.982	0.992	
$\Delta R/R$	0.352	0.234	0.058		0.314	0.248	0.163	

AGA	H_2O				APA	H_2O		
$V_{S,\max}$	164.1	125.9	109.5	103.9	169.4	121.7	102.9	107.9
$V_{S,\min}$	-164.4	-130.4	-101.2	-94.0	-164.4	-130.8	-102.7	-99.8
$\overline{V}_S^+$	68.8	57.8	48.8	46.6	62.2	48.3	40.9	40.0
$\overline{V}_S^-$	-62.3	-53.4	-39.3	-36.0	-61.8	-55.7	-42.0	-41.0
$\overline{V}_S$	0.9	-2.0	2.8	4.0	0.0	-4.2	1.5	2.0
Π	65.5	55.3	44.0	41.3	62.0	52.1	41.4	40.4
σ_+^2	1704.8	953.1	574.4	513.0	1796.3	822.8	556.3	549.3
σ_-^2	1556.8	1038.3	746.6	632.7	1553.1	963.8	667.2	586.5
σ_{tot}^2	3261.7	1991.4	1321.0	1145.7	3349.4	1786.6	1223.5	1135.9
ν	0.249	0.250	0.246	0.247	0.249	0.248	0.248	0.250
RMS	76.9	63.8	51.1	47.7	74.3	60.0	48.2	46.9

RMSD	32.2	20.4	6.7	31.0	17.9	4.4
r	0.976	0.980	0.993	0.971	0.981	0.996
Δ*R/R*	0.531	0.369	0.136	0.525	0.337	0.092

AYA	**EtOH**			
$V_{S,\max}$	165	126.6	105.6	120.6
$V_{S,\min}$	-159.7	-124.1	-97.1	-91.5
$\overline{V}_S^+$	56.1	46.9	35.5	44.6
$\overline{V}_S^-$	-47.8	-44	-33.8	-32
$\overline{V}_S$	-1.1	-2.9	1.7	-0.7
Π	51.4	45.1	34.6	37
σ_+^2	1840.6	1143.2	693.7	1013.7
σ_-^2	1419.1	884.8	605.6	540.5
σ_{tot}^2	3259.7	2028	1299.3	1554.2
ν	0.246	0.246	0.249	0.227
RMS	65.4	55.3	43.0	46.4
RMSD	24.3	14.8	9.7	
r	0.963	0.974	0.981	
Δ*R/R*	0.441	0.292	0.216	
all	peptides			
<*RMS*>	71.8	61.1	48.7	49.8
<*RMSD*>	25.7	16.3	6.5	
<*r*>	0.976	0.982	0.992	
<Δ*R/R*>	0.433	0.296	0.133	

Table 6S. Computed Electrostatic Potential surface quantities for the aromatic molecules.

Surface quantity	ELMAM	ELMAM2	THEO	ELMAM	ELMAM2	THEO
benzene				**catechol**		
$V_{S,\max}$	39.3	20.5	18.9	62.7	62.9	60.2
$V_{S,\min}$	-56.2	-36.5	-27.2	-59.6	-45.0	-36.3
$\overline{V}_S^+$	16.2	10.9	9.4	23.5	18.7	16.1
$\overline{V}_S^-$	-24.9	-15.1	-13.3	-21.9	-18.3	-15.1
$\overline{V}_S$	-5.2	0.2	0.3	-2.2	0.7	1.0
Π	20.8	12.6	10.9	22.4	18.5	15.6
σ_+^2	87.5	21.3	15.6	209.5	122.7	103.8
σ_-^2	212.2	94.6	66.9	156.6	108.2	68.1
σ_{tot}^2	299.7	115.9	82.5	366.0	230.9	171.9
ν	0.207	0.150	0.153	0.245	0.249	0.239
RMS	24.5	14.7	12.6	26.3	21.4	18.2
RMSD	13.5	2.4		11.7	6.0	
r	0.961	0.995		0.935	0.967	
$\Delta R/R$	0.767	0.178		0.536	0.305	

resorcinol				**dihydroxybenzoic acid**		
$V_{S,\max}$	58.7	61.0	55.4	67.9	68.0	63.8
$V_{S,\min}$	-58.1	-42.3	-37.4	-55.1	-38.4	-35.9
$\overline{V}_S^+$	26.3	22.8	21.0	23.6	16.4	18.2
$\overline{V}_S^-$	-23.9	-19.3	-19.0	-17.6	-12.2	-14.0
$\overline{V}_S$	-1.5	1.3	0.4	0.7	2.8	2.3
Π	24.9	21.1	20.0	20.4	14.4	16.2
σ_+^2	223.9	123.4	112.7	254.8	158.9	157.9
σ_-^2	174.5	104.9	87.9	134.5	76.5	75.5
σ_{tot}^2	398.4	228.3	200.6	389.3	235.4	233.3
ν	0.246	0.248	0.246	0.226	0.219	0.219
RMS	28.7	23.7	22.3	24.7	18.2	19.6
RMSD	10.0	5.1		9.2	5.7	

r	0.955	0.978		0.944	0.957
Δ*R/R*	0.400	0.222		0.417	0.302

quercetin			**nitroaniline**	
$V_{S,\max}$	72.8	76.9	80.2	68.4
$V_{S,\min}$	-60.4	-44.9	-50.5	-51.2
$\overline{V}_S^+$	18.4	19.6	32.3	24.7
$\overline{V}_S^-$	-14.2	-13.7	-25.9	-22.1
$\overline{V}_S$	3.4	2.8	1.3	3.3
Π	16.5	16.8	29.0	23.3
σ_+^2	197.1	208.7	434.4	222.9
σ_-^2	158.2	113.0	157.9	230.2
σ_{tot}^2	355.2	321.7	592.4	453.0
ν	0.247	0.228	0.196	0.250
RMS	21.3	21.1	33.7	27.9
RMSD	6.7		10.3	
r	0.950		0.964	
Δ*R/R*	0.316		0.335	

nitrophenol	**alpha**			**Nitrophenol**	**beta**	
$V_{S,\max}$	97.5	76.4	71.7	101.8	78.4	77.5
$V_{S,\min}$	-58.1	-32.2	-42.0	-57.6	-32.1	-41.2
$\overline{V}_S^+$	29.3	20.3	18.2	29.1	20.5	17.8
$\overline{V}_S^-$	-26.5	-12.8	-18.6	-26.7	-12.8	-17.9
$\overline{V}_S$	-1.3	3.5	4.5	-1.3	3.4	4.4
Π	27.7	16.8	17.7	27.7	16.8	17.2
σ_+^2	477.3	209.1	178.1	493.3	214.0	196.2
σ_-^2	183.8	69.8	156.1	181.8	69.2	154.2
σ_{tot}^2	661.1	279.0	334.2	675.1	283.2	350.4
ν	0.201	0.188	0.249	0.197	0.185	0.246
RMS	33.0	20.6	22.5	33.1	20.7	22.3
RMSD	21.6	6.0		21	5.8	
r	0.786	0.964		0.806	0.966	

Δ*R/R*	0.791	0.280		0.772	0.271	

***p*-nitrobenzoic**	**acid**			**dinitrobenzene**		
$V_{S,\max}$	56.6	48.6	47.7	59.2	48.1	49.5
$V_{S,\min}$	-36.2	-32.0	-37.4	-37.2	-22.4	-30.7
$\overline{V}_S^+$	23.2	15.3	16.3	23.0	16.5	22.2
$\overline{V}_S^-$	-13.2	-13.4	-18.2	-12.0	-10.9	-16.1
$\overline{V}_S$	0.6	4.2	5.4	0.9	5.9	7.7
Π	17.1	14.1	15.3	16.3	13.4	18.4
σ_+^2	207.1	117.6	99.4	258.5	103.8	91.0
σ_-^2	57.2	72.3	84.2	90.8	30.0	56.1
σ_{tot}^2	264.2	189.9	183.6	349.3	133.9	147.1
ν	0.17	0.236	0.248	0.192	0.174	0.236
RMS	20.6	17.7	19.5	20.9	15.9	21.9
RMSD	17.2	8.2		24.1	6.7	
r	0.652	0.900		0.376	0.969	
Δ*R/R*	0.857	0.444		1.126	0.358	

All aromatic			
<*RMS*>	26.5	20.8	20.8
<*RMSD*>	16.0	6.3	
<*r*>	0.802	0.961	
<Δ*R/R*>	0.708	0.301	

Supplementary Figures

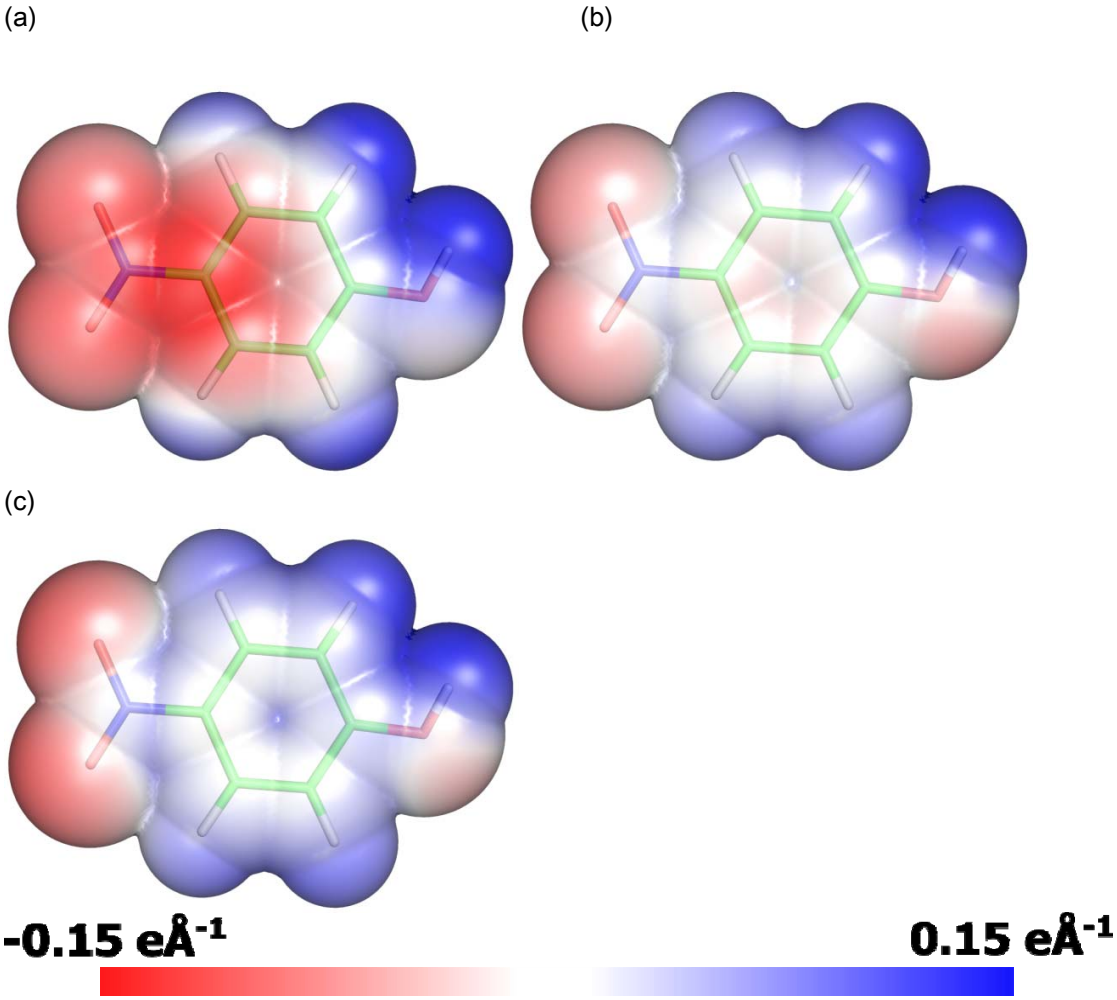


Figure 1S. The electrostatic potential mapped the van der Waals surface in the nitrophenol (alpha) molecule for (a) ELMAM (b) ELMAM2 and (c) THEO_CRYS models. The maximum negative (blue) and positive (red) values of the ESP correspond to -0.15 and 0.15 eÅ^{-1} values, respectively. The $\Delta V/V$ values are 0.791 and 0.280 for ELMAM and ELMAM2, respectively.

References of molecules used for the ELMAM2 Database construction

Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C. & Khmou, A. (2000). *Acta Cryst.* B**56**, 155-165.

Birkedal, H., Madsen, D., Mathiesen, R. H., Knudsen, K., Weber, H.-P., Pattisond, P. & Schwarzenbach, D. (2004). *Acta Cryst.* A**60**, 371-381.

Bouhmaida, N., Bonhomme, F., Guillot, B., Jelsch, C. & Ghermani, N. E. (2009). *Acta Cryst.* B**65**, 363-374.

Chen, Y-S., Stash, A. I. & Pinkerton, A. A. (2007). *Acta Cryst.* B**63**, 309-318.

Coppens, P., Abramov, Y., Carducci, M., Korjov, B., Novozhilova, I., Alhambra, C. & Pressprich, M. R. (1999). *J. Am. Chem. Soc.* **121**, 2585-2593.

Dahaoui, S., Jelsch, C., Howard, J. A. K. & Lecomte, C. (1999). *Acta Cryst.* B**55**, 226-230.

Destro, R., Marsh, R. E. & Bianchi, R. (1988). *J. Phys. Chem.* **92**, 966-973.

Dittrich, B., Koritsánszky, T., Grosche, M., Scherer, W., Flaig, R., Wagner, A., Krane, H. G., Kessler, H., Riemer, C., Schreurs, A. M. M., & Luger, P. (2002). *Acta Cryst.* B**58**, 721-727.

Dittrich, B., Munshi, P. & Spackman, M. A. (2007). *Acta Cryst.* B**63**, 505-509.

Dittrich, B. & Spackman, M. A. (2007). *Acta Cryst.* A**63**, 426-436.

Domagała, S., Korybut-Daszkiewicz, B., Straver, L., & Woźniak, K., (2009). *Inorg. Chem.* **48**, 4010-4020.

Dominiak, P. M., Grech, E., Barr, G., Teat, S., Mallinson, P. & Woźniak, K., (2003). *Chem. Eur. J.* **9**, 963-970.

Espinosa, E., Lecomte, C., Molins, E., Veintemillas, S., Cousson, A. & Paulus, W. (1996). *Acta Cryst.* B**52**, 519-534.

Fournier, B., Bendeif, E-E., Guillot, B., Podjarny, A., Lecomte, C. & Jelsch, C. (2009). *J. Am. Chem. Soc.* **131**, 10929-10941.

Ghermani, N. E., Spasojević-de Biré, A., Bouhmaida, N., Ouharzoune, S., Bouligand, J., Layre, A., Gref, R. & Couvreur, P. (2004). *Pharmaceutical Research*, **21**, 598-607.

Guillot, B., Muzet, N., Artacho, E., Lecomte, C. & Jelsch, C. (2003). *J. Phys. Chem. B*, **107**, 9109-9121.

Howard, J. A. K., Mahon, M. F., Raithby, P. R. & Sparkes, H. A. (2009). *Acta Cryst.* B**65**, 230-237.

Hübschle, C. B., Dittrich, B., Grabowsky, S., Messerschmidt, M. & Luger, P. (2008). *Acta Cryst.* B**64**, 363-374.

Kalinowski, R., Dittrich, B., Hübschle, C. B., Paulmann, C. & Luger, P. (2007). *Acta Cryst.* B**63**, 753-767.

Luger, P., Messerschmidt, M., Scheins, S. & Wagner, A. (2004). *Acta Cryst.* A**60**, 390-396.

Lutz, M., Spek, A. L., van der Geer, E. P. L., van Koten, G. & Gebbink, R. J. M. K. (2008). *Acta Cryst.* C**64**, o87-o90.

Madsen, A. Ø., Sørensen, H. O., Flensburg, C., Stewart, R. F. & Larsen, S. (2004). *Acta Cryst.* A**60**, 550-561.

Martin, A. & Pinkerton, A. A. (1998). *Acta Cryst.* B**54**, 471-477.

Meents, A., Dittrich, B., Johnas, S. K. J., Thome, V. & Weckert, E. F. (2008). *Acta Cryst.* B**64**, 42-49.

Munshi, P. & Guru Row, T. N. (2002). *Acta Cryst.* B**58**, 1011-1017.

Munshi, P. & Guru Row, T. N. (2005). *J. Phys. Chem. A*, **109**, 659-672.

Munshi, P. & Guru Row, T. N. (2006a). *Cryst. Growth & Des*. **6**, 708-718.

Munshi, P. & Guru Row, T. N. (2006b). *Acta Cryst.* B**62**, 612-626.
Munshi, P., Thakur, T. S., Guru Row, T. N. & Desiraju, G. R. (2006c). *Acta Cryst.* B**62**, 118-127.
Ogawa, K., Noda, Y., Lüthi, T. & Buergi, H.-B. (2006). *Priv. Comm.*
Overgaard, J. & Hibbs, D. E. (2004). *Acta Cryst.* A**60**, 480-487.
Pichon-Pesme, V., Lachekar, H., Souhassou, M. & Lecomte, C. (2000). *Acta Cryst.* B**56**, 728-737.
Parrish, D., Zhurova, E. A., Kirschbaum, K. & Pinkerton, A. A. (2006). *J. Phys. Chem. B*, **110**, 26442-26447.
Rodrigues, B. L., Tellgren, R. & Fernandes, N. G. (2001). *Acta Cryst.* B**57**, 353-358.
Scheins, S., Dittrich, B., Messerschmidt, M., Paulmann, C. & Luger, P. (2004). *Acta Cryst.* B**60**, 184-190.
Slouf, M., Holy, A., Petricek, V. & Cisarova, I. (2002). *Acta Cryst.* B**58**, 519-529.
Sørensen, H. O., Stewart, R. F., McIntyre, G. J. & Larsen, S. (2003). *Acta Cryst.* A**59**, 540-550.
Sparkes, H. A., Brayshaw, S. K., Weller, A. S. & Howard, J. A. K. (2008). *Acta Cryst.* B**64**, 550-557.
Volkov, A., Abramov, Y., Coppens, P. & Gatti C. (2000). *Acta Cryst.* A**56**, 332-339.
Wiest, R., Pichon-Pesme, V., Bénard, M. & Lecomte, C. (1994). *J. Phys. Chem.* **98**, 1351-1362.
Zhurov, V. V., Zhurova, E. A., Chen Y-S. & Pinkerton, A. A. (2005). *J. Appl. Cryst.* **38**, 827-829.
Zhurova, E. A., Martin, A. & Pinkerton, A. A. (2002) *J. Am. Chem. Soc.* **124**, 8741-8750.
Zhurova, E. A., Matta, C. F., Wu, N., Zhurov, V. V. & Pinkerton, A. A. (2006). *J. Am. Chem. Soc.* **128**, 8849-8861.
Zhurova, E. A. & Pinkerton, A. A. (2001). *Acta Cryst.* B**57**, 359-365.