

QSPR Modeling of the Mineral Crystal Lattice Energy by Optimization of Correlation Weights of Vertex and Vertex Degree in Graph of Atomic Orbitals: Linear and Non-Linear Models

¹E.A. Castro, ²A.P. Toropova, ²A.A. Toropov and ²R.Z. Akhmerov

¹INIFTA, Theoretical Chemistry Division, Department of Chemistry,
Faculty of Exact Sciences, La Plata National University,
Diagonal 113 and 64, Sucursal 4, C.C. 16, La Plata 1900, Buenos Aires, Argentina
²Uzbekistan Academy of Sciences Institute of Geology and Geophysics 700041,
Khodzhibaev Street 49, Tashkent, Uzbekistan

Abstract: Models of the crystal lattice of minerals of general formula of A_mC_n : $m, n = 1, 2$; $A = \text{Li, K, Na, Mg, Ca, Mn, Cu, Zn, Sr, Cd, Ba, Hg, Pb, Cs}$ and Rb ; $C = \text{Be, O, F, S, Cl, Br}$ and I ; as a mathematical function of their structure have been constructed. As the elucidations of molecular structure the Graph of Atomic Orbitals (GAOs) have been used (GAO vertexes are atomic orbitals, i.e., $1s^2, 2p^5, 3d^{10}$, etc). Optimization of Correlation Weights of Local Invariants (OCWLI) has been used as method of modeling. Comparison of three versions OCWLI based on descriptors with exponent of one (DCW¹), two (DCW²) and three (DCW³) has shown that best statistical characteristics take place for case of the exponent two.

Key words: QSPR modeling % crystal lattice energy % flexible topological descriptors

INTRODUCTION

Lattice Potential Energy (U , kJ mol^{-1}) of ionic minerals is an important parameter widely used in inorganic chemistry, geochemistry (mineralogy, crystallography) and chemical technology [1-3]. Kapustinski equation has, more than 40 years, been enormously powerful and useful in providing approximate values of lattice energy for inorganic salts (minerals of different ionic architecture). As alternative of Kapustinski equation correlation weighting of molecular invariants can be used [4-6]. In Ref. [4] comparison between predicting crystal lattice energy based on Kapustinski calculations and predicting based on optimal descriptors calculated with special molecular graph representing architecture of ionic crystal lattice has been reported. In Ref. [7] so-called Graph of Atomic Orbitals (GAO) has been examined as an alternative of molecular graph traditionally used in Quantitative Structure-property Relationships (QSPR). Main idea of the GAO is to take into account structures of chemical elements in QSPR analysis. The aim of the present study is to compare

predictive ability of three modes of modeling crystal lattice energy of ionic minerals: linear model, i.e., model defined as $U = F(\text{descriptor}^1)$, quadratic, i.e., $U = F(\text{descriptor}^2)$ and cubic $U = F(\text{descriptor}^3)$.

MATERIALS AND METHODS

Optimal descriptors have been calculated as:

$$\text{DCW} = \prod_{k=1}^N \text{CW}(V_k) \times \text{CW}(\text{VD}_k) \quad (1)$$

Where, V_k is k -th vertex in Graph of Atomic Orbitals (GAO), VD_k is number of neighboring vertexes of the V_k (vertex degree); $\text{CW}(V_k)$ and $\text{CW}(\text{VD}_k)$ are the Correlation Weights (CWs) of mentioned GAO invariants. By means of Monte Carlo method optimization procedure Cws which are producing maximal values of the correlation coefficient between the descriptor in degree of m and lattice energy values can be calculated. Having numerical data on the CWs one can calculate descriptors values for both the training and test sets. Model should be defined with structures of the training set (solely):

Corresponding Author: Dr. E.A. Castro, INIFTA, Theoretical Chemistry Division, Department of Chemistry, Faculty of Exact Sciences, La Plata National University, Diagonal 113 and 64, Sucursal 4, C.C. 16, La Plata 1900, Buenos Aires, Argentina

Table 1: Statistical characteristics of lattice potential energy models

Probe No.	C ₁	C ₀	Training set, n = 65			Test set, n = 21		
			r ²	s kJ molG ¹	F	r ²	s kJ molG ¹	F
U = F (DCW ¹)								
1	101.722	-3054.873	0.9948	78.058	12076	0.9878	99.886	1542
2	104.037	-3030.510	0.9949	77.399	12284	0.9895	93.628	1784
3	101.718	-3047.272	0.9948	78.261	12013	0.9896	93.095	1804
U = F (DCW ²)								
1	0.997	-761.447	0.9957	71.264	14501	0.9919	84.029	2325
2	1.144	-764.602	0.9957	71.297	14488	0.9918	84.219	2286
3	1.085	-756.111	0.9957	71.058	14586	0.9928	80.382	2613
U = F (DCW ³)								
1	0.012	-16.342	0.9966	63.327	18381	0.9939	90.248	3092
2	0.013	-5.170	0.9965	64.122	17927	0.9929	97.206	2649
3	0.012	-24.635	0.9966	63.404	18336	0.9943	85.607	3297

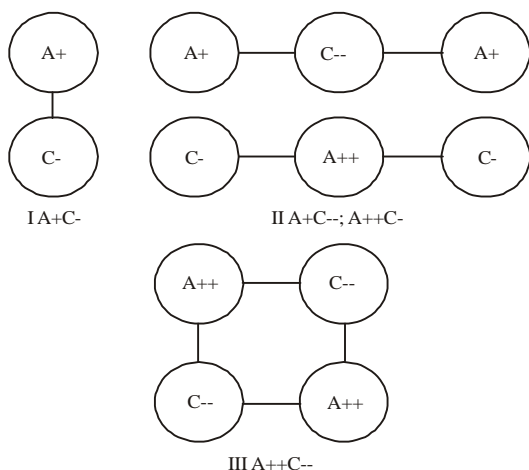


Fig. 1: General scheme of ionic crystal mineral under consideration

$$U(\text{pot}) = C_0 + C_1 * \text{DCW}^m \quad (2)$$

Predictive potential of Eq. (2) can be estimated with structures of an external test set. Results of these actions are shown in Table 1. Scheme of three kinds of minerals under consideration are shown in Fig. 1.

GAO can be obtained from molecular graph by replacing each atom by group of atomic orbitals. These groups for atoms under consideration are the following:

Li (1s² 2s¹), Be (1s² 2s²), O (1s² 2s² 2p⁴), F(1s² 2s² 2p⁵), Na(1s² 2s² 2p⁶ 3s¹), Mg(1s² 2s² 2p⁶ 3s²), S(1s² 2s² 2p⁶ 3s² 3p⁴), Cl(1s² 2s² 2p⁶ 3s² 3p⁵), K(1s² 2s² 2p⁶ 3s² 3p⁶ 4s¹), Ca(1s² 2s² 2p⁶ 3s² 3p⁶ 4s²), Mn(1s² 2s² 2p⁶ 3s² 3p⁶ 3d⁵ 4s²), Cu(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s¹), Zn(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰

4s²), Br(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁵), Sr(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 5s²), Cd(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 5s²), J(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 5s² 5p⁵), Ba(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 5s² 5p⁶ 6s²), Hg(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 5s² 5p⁶ 5d¹⁰ 6s²), Pb(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 4f¹⁴ 5s² 5p⁶ 5d¹⁰ 6s² 6p²), Cs(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 4f¹⁴ 5s² 5p⁶ 5d¹⁰ 6s² 6p¹), Rb(1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p⁶ 4d¹⁰ 5s² 5p⁶ 5s¹)

RESULTS

Eighty six ionic minerals under consideration have been randomly split into training (n = 65) and test (n = 21) set. From Table 1, one can see that best statistical characteristics take place for case of the m = 2 (quadratic model). Correlation weights obtained in three probes of the Monte Carlo optimization with DCW² are shown in Table 2. Experimental data on the U (kJ molG¹) taken from Ref. 4, training and test sets and calculated values of the lattice potential energy are present in Table 3.

Table 2: Correlation weights for the quadratic model U = F (DCW²)

GAO invariant	CWs of probe 1	CWs of probe 2	CWs of probe 3
V _k			
1s2	7.56505	6.57634	6.23653
2s1	3.70293	4.78467	4.40878
2s2	2.03727	3.11076	2.71529
2p4	4.15744	3.92932	4.07607
2p5	4.45851	4.33092	4.33234
2p6	0.22099	0.19672	0.18558
3s1	1.72488	1.71777	1.73626
3s2	0.27361	0.22844	0.22728
3p4	2.31300	2.20687	2.37046

Table 2: Continue

GAO invariant	CWs of probe 1	CWs of probe 2	CWs of probe 3
3p5	3.12811	3.04162	3.10062
3p6	0.42543	0.42052	0.68636
3d5	0.11143	0.22844	0.21505
3d10	0.47839	0.56860	0.60204
4s1	1.12579	1.11613	0.95026
4s2	0.53593	0.49636	0.42635
4p5	2.05503	1.88818	1.59658
4p6	0.33477	0.40847	0.27214
4d10	0.52494	0.27515	0.31988
4f14	1.10620	1.61256	1.09892
5s1	0.52750	0.52970	0.40348
5s2	0.40268	0.51395	0.32057
5p5	0.04249	0.03715	0.05833
5p6	0.46169	0.31734	0.48915
5d10	1.54073	1.41734	1.81960
6s1	0.33062	0.43209	0.35253
6s2	1.24458	1.20078	1.14584
6p2	0.05079	0.25066	0.21176
V_{d_k}			
0002	1.78280	1.66555	1.81940
0003	1.66677	1.55683	1.68879
0004	1.61591	1.51157	1.64836
0005	1.54760	1.44913	1.56798
0006	1.47121	1.38141	1.49553
0007	1.55334	1.44959	1.56862
0008	1.40911	1.32018	1.44381
0009	1.36181	1.26402	1.37329
0010	1.44363	1.35250	1.46650
0011	1.53865	1.42953	1.55980
0012	1.20009	1.13111	1.21478
0014	1.28548	1.19665	1.28191
0015	1.13249	1.03974	1.11493
0016	1.35549	1.28088	1.41463
0018	1.00697	0.91645	1.00577
0020	1.09468	1.02638	1.11691
0022	1.42309	1.32524	1.46487
0024	0.70577	0.67862	0.72054
0028	0.65400	0.58700	0.61621
0030	0.52629	0.44725	0.46817

Table 3: Calculation crystal lattice energy by means of equation $U = -761.447 + 0.997 DCW^2$ (first probe DCW^2 -based optimization). Statistical characteristics of this model on the test set are the following: $n = 21$, $R^2 = 0.9919$, $s = 84$ kJ mol G^1

n	Structure	DCW	U_{expr}	U_{calc}	$U_{\text{expr}} - U_{\text{calc}}$
Training set					
1	Li $^+$ Cl 1	1682.30	844.36	916.14	-71.78
2	Cd $^{++}$ I 1	3057.11	2353.34	2287.10	66.24
3	Li $^+$ F 1	1618.56	739.86	852.58	-112.72
4	Na $^+$ F 1	1761.41	915.42	995.03	-79.61
5	Na $^+$ Cl 1	1539.97	785.84	774.21	11.63
6	Rb $^+$ I 1	1337.03	622.82	571.83	50.99
7	Cs $^+$ F 1	1562.46	748.22	796.63	-48.41
8	Cs $^+$ Cl 1	1362.00	652.08	596.73	55.35
9	Be $^{++}$ F 1	4129.65	3452.68	3356.63	96.05

Table 3: Continue

n	Structure	DCW	U_{expr}	U_{calc}	$U_{\text{expr}} - U_{\text{calc}}$
10	Mg $^{++}$ F 1	3635.70	2880.02	2864.07	15.95
11	Mg $^{++}$ Cl 1	3285.25	2487.10	2514.60	-27.50
12	Ca $^{++}$ Cl 1	3011.56	2194.50	2241.67	-47.17
13	Ca $^{++}$ I 1	2864.62	2035.66	2095.15	-59.49
14	Ba $^{++}$ Br 1	2733.65	1935.34	1964.55	-29.21
15	Ba $^{++}$ I 1	2648.59	1839.20	1879.72	-40.52
16	Zn $^{++}$ Cl 1	3333.98	2683.56	2563.19	120.37
17	Cd $^{++}$ Br 1	3202.63	2478.74	2432.21	46.53
18	Pb $^{++}$ I 1	2869.46	2077.46	2099.98	-22.52
19	Hg $^{++}$ I 1	3297.55	2633.40	2526.87	106.53
20	Mn $^{++}$ Cl 1	3273.08	2462.02	2502.47	-40.45
21	Cu $^{++}$ Cl 1	3433.04	2758.80	2661.98	96.82
22	Li $^+$ O $^-$	3575.66	2938.54	2804.19	134.35
23	Ba $^{++}$ O $^-$	3856.81	3122.46	3084.56	37.90
24	Sr $^{++}$ O $^-$	4073.88	3306.38	3301.02	5.36
25	Na $^+$ O $^-$	3350.72	2604.14	2579.88	24.26
26	K $^+$ O $^-$	3062.79	2353.34	2292.77	60.57
27	Rb $^+$ O $^-$	3045.72	2273.92	2275.74	-1.82
28	Cs $^+$ O $^-$	3027.05	2202.86	2257.12	-54.26
29	Be $^{++}$ O $^-$	5288.65	4514.40	4512.39	2.01
30	Mg $^{++}$ O $^-$	4690.53	3929.20	3915.94	13.26
31	Hg $^{++}$ O $^-$	4687.72	3929.20	3913.14	16.06
32	Pb $^{++}$ O $^-$	4238.25	3553.00	3464.93	88.07
33	Mn $^{++}$ O $^-$	4656.94	3845.60	3882.45	-36.85
34	Cd $^{++}$ O $^-$	4594.84	3807.98	3820.53	-12.55
35	Cu $^{++}$ O $^-$	5049.12	4138.20	4273.53	-135.33
36	Mg $^{++}$ S $^-$	4112.18	3252.04	3339.22	-87.18
37	Ca $^{++}$ S $^-$	3808.64	3017.96	3036.52	-18.56
38	Pb $^{++}$ S $^-$	3935.32	3059.76	3162.85	-103.09
39	Mn $^{++}$ S $^-$	4115.90	3515.38	3342.92	172.46
40	Cu $^{++}$ S $^-$	4478.00	3720.20	3704.01	16.19
41	Sr $^{++}$ S $^-$	3651.96	2871.66	2880.28	-8.62
42	Ba $^{++}$ S $^-$	3528.09	2742.08	2756.76	-14.68
43	Cd $^{++}$ S $^-$	4111.89	3352.36	3338.92	13.44
44	K $^+$ F 1	1608.11	810.92	842.15	-31.23
45	K $^+$ I 1	1355.35	639.54	590.10	49.44
46	Pb $^{++}$ F 1	3266.90	2466.20	2496.30	-30.10
47	Rb $^+$ F 1	1585.18	777.48	819.29	-41.81
48	Pb $^{++}$ Br 1	2927.19	2207.04	2157.54	49.50
49	Li $^+$ Br 1	1608.74	798.38	842.79	-44.41
50	Na $^+$ Br 1	1461.96	739.86	696.42	43.44
51	Na $^+$ I 1	1483.75	689.70	718.14	-28.44
52	K $^+$ Br 1	1325.67	677.16	560.51	116.65
53	Cs $^+$ Br 1	1257.62	631.18	492.65	138.53
54	Cs $^+$ I 1	1319.39	601.92	554.24	47.68
55	Be $^{++}$ I 1	3534.64	2800.60	2763.29	37.31
56	Mg $^{++}$ I 1	3121.50	2311.54	2351.31	-39.77
57	Hg $^{++}$ Cl 1	3453.04	2608.32	2681.92	-73.60
58	Ba $^{++}$ F 1	3034.79	2286.46	2264.84	21.62
59	Zn $^{++}$ Br 1	3348.12	2645.94	2577.29	68.65
60	Pb $^{++}$ Cl 1	3001.86	2232.12	2232.01	0.11
61	Mn $^{++}$ I 1	3112.53	2265.56	2342.37	-76.81
62	Ba $^{++}$ Cl 1	2775.48	1956.24	2006.25	-50.01
63	Li $^+$ F 1	1922.72	1032.46	1155.88	-123.42
64	Mn $^{++}$ Br 1	3290.80	2319.90	2520.14	-200.24
65	Hg $^{++}$ S $^-$	4323.75	3519.56	3550.19	-30.63

Table 3: Continue

n	Structure	DCW	U_{expr}	U_{calc}	$U_{\text{expr}}-U_{\text{calc}}$
Test set					
1	Be ⁺⁺ Br ^l	3779.60	2892.56	3007.56	-115.00
2	Sr ⁺⁺ Cl ^l	2887.91	2106.72	2118.37	-11.65
3	Sr ⁺⁺ Br ^l	2878.97	2044.02	2109.46	-65.44
4	Rb ⁺ Cl ^l	1383.73	685.52	618.41	67.11
5	Rb ⁺ Br ^l	1294.62	660.44	529.54	130.90
6	Be ⁺⁺ Cl ^l	3723.12	2980.34	2951.24	29.10
7	Mg ⁺⁺ Br ^l	3319.17	2411.86	2548.42	-136.56
8	Ca ⁺⁺ F ^l	3322.59	2579.06	2551.84	27.22
9	Sr ⁺⁺ F ^l	3175.79	2424.40	2405.45	18.95
10	Zn ⁺⁺ F ^l	3679.07	3001.24	2907.32	93.92
11	Sr ⁺⁺ I ^l	2750.23	1952.06	1981.08	-29.02
12	Zn ⁺⁺ I ^l	3171.08	2591.60	2400.75	190.85
13	Cd ⁺⁺ F ^l	3531.61	2767.16	2760.27	6.89
14	Cd ⁺⁺ Cl ^l	3210.49	2499.64	2440.05	59.59
15	Hg ⁺⁺ Br ^l	3386.48	2608.32	2615.54	-7.22
16	Ca ⁺⁺ Br ^l	3023.01	2123.44	2253.09	-129.65
17	K ⁺ Cl ^l	1405.01	706.42	639.63	66.79
18	Cu ⁺⁺ Br ^l	3441.28	2725.36	2670.20	55.16
19	Ca ⁺⁺ O ⁻	4298.64	3519.56	3525.15	-5.59
20	Zn ⁺⁺ O ⁻	4805.47	4054.60	4030.57	24.03
21	Zn ⁺⁺ S ⁻	4252.97	3561.36	3479.61	81.75

CONCLUSIONS

Quadratic model of the lattice potential energy is better than both linear and cubic models.

REFERENCES

1. Jenkins, H.D.B. and H.K. Roobottom, 1999. J. Passmore, L. Glasser, Inorg. Chem., 38: 3609-3620.
2. Jenkins, H.D.B. and D. Tudela, 2002. L. Glasser, Inorg. Chem., 41: 2364-2367.
3. Jenkins, H.D.B., 2002. L. Glasser, Inorg. Chem., 41: 4378-4388.
4. Toropova, A.P. and A.A. Toropov, 2001. Russ. J. Struct. Chem., 42: 1230-1232.
5. Castro, E.A., A.P. Toropova, A.A. Toropov and D.V. Mukhamedzhanova, 2003. J. Argent. Chem. Soc., 91: 85-90.
6. Toropov, A.A., A.P. Toropova and D.V. Mukhamedzhanova, 2005. I. Gutman, Ind. J. Chem., Sect A, 44: 1545-1552.
7. Toropov, A.A., A.P. Toropova, A.I. Nesterove and O.M. Nabiev, 2004. Chem. Phys. Lett., 384: 357-363.