



## Calculation of Gas-phase Gibb's Free Energy Changes of Some Small Molecules with Monte Carlo, DFT (MPW1PW91), Composite (CBS-QB3), Gaussian-n (G1, G2) and Gaussian modified (G2MP2) methods

MUHAMMAD ABUL KASHEM LITON, MD. NURUZZAMAN\* and SABRINA HELEN

Department of Chemistry, Faculty of Science, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh.

\*Corresponding author E-mail: md.nuruzzaman04@gmail.com

<http://dx.doi.org/10.13005/ojc/350305>

(Received: April 17, 2019; Accepted: June 04, 2019)

### ABSTRACT

A set of 74 small molecules was employed to generate a quantitative structure-property relationship (QSPR) model for predicting gas-phase Gibb's free energy changes ( $\Delta G_{\text{gas}}$ ) on the simplified molecular input line entry system (SMILES). To obtain the models, the Monte Carlo method (MCM) was applied to calculate the descriptors. The best proposed model of them provides an excellent statistical result of  $r^2 = 0.9866$ ,  $q^2 = 0.9857$ ,  $s = 2.48$  and  $F = 4346$  for the training set, and  $r^2 = 0.9340$ ,  $q^2 = 0.8418$ ,  $s = 5.80$ ,  $F = 149$  for the test set. Consequently, the gas-phase Gibb's free energy changes were also calculated in DFT (MPW1PW91/6-311+G(2d,2p)), CBS-QB3, G1, G2 and G2MP2 methods which showed a very good agreement with experimental values.

**Keywords:** QSPR, SMILES, Monte Carlo, G1, G2, CBS-QB3, Training set, Test set.

### INTRODUCTION

The Gibb's free energy or chemical potential in basic theoretical research plays an important role on chemical and biological systems<sup>1</sup>. These are the central quantities that determine the route of spontaneous change for a system or explain the ability of biological systems in all organisms to resist a natural tendency to disorder. Examination of free energy behavior of such systems enables us to fully understand the structural, kinetic and dynamic characteristics of processes like – protein

ligand binding constants<sup>2</sup>, permeability coefficients<sup>3</sup>, conformational equilibrium constants<sup>4</sup>, membrane-water partition coefficient<sup>5,6</sup>, electron transfer<sup>7</sup>, solvation<sup>8</sup> etc.

Numerous reports have been appeared in the literature describing computational methods for the calculation of gas-phase Gibbs free energy<sup>9-14</sup> but calculation in computational methodology along with QSPR modeling is not randomly found. In recent years, rapid advances in theory of calculations, continuous increasing of computational



capabilities as well as statistical mechanics make the determination of Gibb's free energy changes more reliable, efficient and well organized<sup>15-18</sup>.

For an instance, protein–ligand binding constants and membrane–water partition coefficients are an essential part of rational drug design and cannot be evaluated reliably and accurately without the knowledge of the associated free energy changes. The ability to calculate free energy in an efficient manner is of paramount importance in the structural and thermodynamic study of gas hydrate systems<sup>19</sup>. Thus MDFE simulations have been improved by new treatments of long-range electrostatic interactions and, successful in determination of protein–ligand binding constants<sup>2</sup>.

In this study we have developed a QSPR model and established a computational methodology for the calculation of gas phase Gibb's free energy of some small molecules which have many applications in nature. In addition we have focused our study to calculate the free energies as an essential complement to experimental quantities from reported literatures and to establish the accurate, low cost-effective and reliable methods for predicting the gas-phase Gibbs free energy changes.

## MATERIALS AND METHODS

### Data

Table 1 shows structural details of the 74 small molecules used in this study. The experimental gas phase Gibb's free energy changes (at 298.15 K and 1 atmospheric pressure) were taken from reported literatures. All the gas-phase Gibb's free energy changes in data set represented in kcal/mol. The data set was splitted into two: a training set (61) and a test set (13) where the molecules in test set were arranged in random manner.

### Descriptors calculation

Three kinds of descriptors: graph-based, SMILES-based, and hybrid descriptors were generated using CORAL software<sup>20</sup>. Furthermore three kinds of molecular graphs: hydrogen suppressed graph (HSG), hydrogen filled graph (HFG), and graph of atomic orbitals (GAO) were also generated by CORAL<sup>21</sup>.

The graph-based optimal descriptors were calculated as the following:

$${}^{\text{GRAPH}}\text{DCW}(\text{Threshold}, N_{\text{epoch}}) = \sum \text{CW}(A_k) + \alpha \sum \text{CW}({}^0\text{EC}_k) + \beta \sum \text{CW}({}^1\text{EC}_k) + \gamma \sum \text{CW}({}^2\text{EC}_k) + \delta \sum \text{CW}({}^3\text{EC}_k) \quad (1)$$

Where,  $A_k$  is a chemical element like C, N, O, etc., for HSG and HFG; or atomic orbitals, such as  $1s^1$ ,  $2p^3$ ,  $3d^{10}$ , etc., for GAO;  ${}^0\text{EC}_k$ ,  ${}^1\text{EC}_k$ , ... ${}^3\text{EC}_k$  represents the hierarchy of the Morgan extended connectivity;  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  can be 1 or 0: combinations of their values gives possibility to define various of the graph-based optimal descriptor;  $\text{CW}(x)$  is the correlation weight of a molecular feature (encoded by  $A^x$  or  ${}^x\text{EC}_k$ ).

Similarly the SMILES-based optimal descriptors were calculated as the following:

$${}^{\text{SMILES}}\text{DCW}(\text{Threshold}, N_{\text{epoch}}) = \alpha \sum \text{CW}(S_k) + \beta \sum \text{CW}(\text{SS}_k) + \gamma \sum \text{CW}(\text{SSS}_k) + x \sum \text{CW}(\text{NOSP}) + y \sum \text{CW}(\text{HALO}) + z \sum \text{CW}(\text{BOND}) \quad (2)$$

Where,  $S_k$ ,  $\text{SS}_k$ , and  $\text{SSS}_k$  are one-, two-, and three-component SMILES attributes, respectively. NOSP, HALO, and BOND are indices that related to the presence or absence of chemical elements like N, O, S and P (NOSP); F, Cl, and Br (HALO). The BOND is related to a mathematical function where double (=), triple (#), or stereo chemical bonds may be present or absent and finally  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ , and  $z$  can be 1 or 0: combinations of their values provide possibility to define various versions of the SMILES-based optimal descriptor.  $\text{CW}(x)$  is the correlation weight of a molecular feature (encoded by  $S_k S_k S_k$  or  ${}^x\text{EC}_k$ ).

The hybrid optimal descriptors were calculated by taking into account both representations of the molecular structure by GRAPH and by SMILES.

$${}^{\text{HYBRID}}\text{DCW}(\text{Threshold}, N_{\text{epoch}}) = {}^{\text{GRAPH}}\text{DCW}(\text{Threshold}, N_{\text{epoch}}) + {}^{\text{SMILES}}\text{DCW}(\text{Threshold}, N_{\text{epoch}}) \quad (3)$$

Threshold and  $N_{\text{epoch}}$  in equations (1) – (3) represent the parameters of the Monte Carlo optimization. The threshold is the criterion for classification of the components of the representation of the molecular structure into two classes: noise (rare) and active (not rare). The correlation weight of a rare component is fixed as zero; hence rare component is not involved in the building up of the model.  $N_{\text{epoch}}$  is the number of epochs of the Monte Carlo optimization.

**Table 1: List of the molecules with experimental values employed for the calculation of gas-phase Gibb's free energies (at 298.15 K and 1 atmospheric pressure)**

ID	Structure	$\Delta G_{\text{gas}}$ (kcal/mol)	Ref	ID	Structure	$\Delta G_{\text{gas}}$ (kcal/mol)	Ref
1		343.2 ± 4.0	22	2		347.0 ± 3.0	23
3		344.4 ± 3.0	24	4		349.0 ± 6.7	25
5		345.0 ± 3.0	26	6		348.9 ± 2.0	27
7		350.6 ± 2.0	27	8		327.5 ± 2.0	28
9		389.1 ± 3.0	29	10		342.3 ± 3.8	30
11		347.1 ± 2.0	27	12		320.9 ± 5.0	31
13		335.5 ± 0.81	32	14		344.1 ± 2.0	33
15		352.8 ± 2.0	34	16		387.9 ± 2.1	35
17		395.9 ± 0.38	36	18		334.7 ± 0.71	37
19		391.7 ± 0.70	38	20		355.7 ± 3.2	39
21		383.7 ± 0.06	40	22		368.5 ± 1.1	41
23		359.2 ± 1.4	42	24		349.1 ± 1.2	43
25		376.2 ± 0.62	44	26		369.5 ± 0.40	45
27		372.6 ± 1.1	41	28		366.6 ± 2.8	46
29		367.6 ± 0.70	47	30		370.0 ± 5.0	48
31		342.5 ± 1.1	49	32		360.0 ± 19	50
33		372.6 ± 2.0	51	34		403.0 ± 12	52
35		352.0 ± 3.0	26	36		377.80 ± 1.1	53
37		401.0 ± 0.50	54	38		364.1 ± 2.0	55
39		372.5 ± 3.0	56	40		372.8 ± 3.0	57
41		357.8 ± 2.0	58	42		391.3 ± 3.0	59

43		404.0 ± 2.1	60	44		382.0 ± 4.0	61
45		376.2 ± 2.0	62	46		383.8 ± 0.10	63
47		379.0 ± 4.0	64	48		358.7 ± 2.0	27
49		361.9 ± 2.0	27	50		398.2 ± 2.2	65
51		367.4 ± 2.0	27	52		407.2 ± 2.1	60
53		411.7 ± 2.1	60	54		401.0 ± 3.0	66
55		366.4 ± 2.0	27	56		344.0 ± 3.0	67
57		349.7 ± 2.0	27	58		383.0 ± 1.7	68
59		359.0 ± 2.0	27	60		365.2 ± 2.0	27
61		396.5 ± 1.6	69	62		394.0 ± 3.0	70
63		408.7 ± 0.8	71	64		400.6 ± 4.1	52
65		381.2 ± 3.6	52	66		370.3 ± 1.5	72
67		368.0 ± 2.0	73	68		349.9 ± 2.0	74
69		389.7 ± 3.4	75	70		361.3 ± 3.0	73
71		342.0 ± 2.0	73	72		381.3 ± 3.2	76
73		356.0 ± 5.1	77	74		356.4 ± 1.6	78

Ref: Reference,  $\Delta G_{\text{gas}}$ : Gibb's free energy changes in gas-phase

### Computational calculation

All the calculations were performed on Intel core-i5 Fujitsu Lap-top Computer with 3–4 GB of memory and 48 GB of scratch disk space, using CORAL<sup>20</sup> and Gaussian 09 software<sup>79</sup>. Every structure is initially optimized at 298.15 K in the gas phase at the Hartree–Fock (HF) level<sup>80–82</sup> and finally G1, G2, G2MP2, CBS-QB3 and DFT(MPW1PW91/6-311+G(2d,2p)) methods were used to calculate the  $G^\circ(\text{g})$  values (Gas phase Gibbs free energy)<sup>83–89</sup>. The optimized structures were confirmed to be energy minima by vibrational frequency calculations with all real frequencies. In this study there are no imaginary frequency found.

These high level composite procedures have been designed particularly for the prediction of reliable energies of molecules in the gas-phase and have been demonstrated to provide the accuracy within 1-2 kcal/mol when assessed against large test sets of thermo-chemical data<sup>90,91</sup>.

### RESULTS AND DISCUSSIONS

The models calculated with GRAPH based, SMILES based and a Hybrid (both GRAPH and SMILES) version of the descriptor. The versions of descriptors  $\text{GRAPHDCW}(\text{Threshold}, N_{\text{epoch}})$  calculated here with Eq. (1) (HSG) are the following:  $\alpha = 0$ ,  $\beta =$

1,  $\gamma = 1$ , and  $\delta = 1$ . The SMILES-based descriptors  $^{SMILES}DCW(Threshold, N_{epoch})$  applied in our work with Eq. (2) are as follows:  $\alpha = 1$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $\delta = 0$ ,  $x = 0$ ,  $y = 1$  and  $z = 0$  and finally the hybrid version of the descriptors  $^{HYBRID}DCW(Threshold, N_{epoch})$  were calculated with Eq. (3).

Equations (4) – (6) represent the models calculated with GRAPH based, SMILES based and Hybrid based descriptors. The classic scheme (only training and test set used) has been applied here is the Monte Carlo optimization method.

The aforementioned models for the calculation of gas-phase Gibb's free energy changes are the following:

$$\Delta G_{gas} = 313.5442(\pm 0.3034) + 6.4783(\pm 0.0409) \times DCW(0, 30) \quad (4)$$

$$\begin{aligned} n=61, r^2=0.8667, q^2=0.8581, s=7.83, \\ MAE=5.01, F=384 \text{ (training set)} \\ n=13, r^2=0.8058, r^2_{pred}=0.7381, \\ s=13.2, MAE=10.08, F=46 \text{ (test set)} \\ R_m^2=0.6632, \Delta R_m^2=0.1896, R_m^{2'}=0.7580, R_m^{2''}=0.5684 \end{aligned}$$

$$\Delta G_{gas} = 313.0736(\pm 0.1995) + 8.5338 (\pm 0.0278) \times DCW(1,30) \quad (5)$$

$$\begin{aligned} n=61, r^2=0.9565, q^2=0.9538, s=4.47, \\ MAE=3.09, F=1296 \text{ (training set)} \\ n=13, r^2=0.8165, r^2_{pred}=0.7183, \\ s=11.0, MAE=8.15, F=49 \text{ (test set)} \\ R_m^2=0.7011, \Delta R_m^2=0.1733, R_m^{2'}=0.7878, R_m^{2''}=0.6145 \end{aligned}$$

$$\Delta G_{gas} = 295.9911(\pm 0.1325) + 4.4749(\pm 0.0085) \times DCW(0,30) \quad (6)$$

$$\begin{aligned} n=61, r^2=0.9866, q^2=0.9857, s=2.48, \\ MAE=1.35, F=4346 \text{ (training set)} \\ n=13, r^2=0.9340, r^2_{pred}=0.8418, \\ s=5.80, MAE=4.83, F=149 \text{ (test set)} \\ R_m^2=0.7382, \Delta R_m^2=0.1081, R_m^{2'}=0.7923, \\ R_m^{2''}=0.6842 \end{aligned}$$

where,  $n$  is the number of compounds in a set;  $r^2$  is squared correlation coefficient;  $q^2$  and  $r^2_{pred}$  are leave-one-out correlation coefficients for training and test sets;  $s$  is the standard error of estimation;  $MAE$  is mean absolute error;  $F$  is Fischer F-ratio; novel validation metrics  $R_m^2$ ;  $R_m^{2'}$  is

mean of  $R_m^{2'}$  and  $R_m^{2''}$  values and  $\Delta R_m^2$  is the absolute difference between  $R_m^{2'}$  and  $R_m^{2''}$ .  $R_m^2$  should be more than 0.5 and  $\Delta R_m^2$  should be smaller than 0.2 .

Table 3 contains the statistical quality of all models for gas phase Gibb's free energy changes of data set calculated by equations (4) – (6) of QSPR methods. In this table the statistics obtained from the HYBRID method are  $n = 61$ ,  $r^2 = 0.9866$ ,  $q^2 = 0.9857$ ,  $s = 2.48$ ,  $MAE = 1.35$  and  $F = 4356$  for the training set and  $r^2 = 0.9340$ ,  $r^2_{pred} = 0.8418$ ,  $s = 5.80$ ,  $MAE = 4.83$  and  $F = 149$  for the test set. The large value of  $F$  in hybrid model indicates that this linear model is robust and hybrid version of optimal descriptors is more compatible with the data than other models. On the basis of other statistical values hybrid method appeared to be a significant improvement among all the QSPR methods.

Table 2 shows the calculated gas-phase Gibb's free energy changes in kcal/mol of all (74) molecules using the three aforementioned QSPR methods of Coral and five methods: DFT (MPW1PW91/6-311+G(2d,2p), Gaussian-n (G1, G2, G2MP2) and composite (CBS-QB3) of Gaussian 09 respectively. The RMSE (Root Mean Squared Error) of gas-phase Gibb's free energy changes calculated with GRAPH, SMILES and Hybrid methods are 8.3, 5.98 and 3.23 respectively. In the case of QSPR analysis the HYBRID model gives the lowest RMSE value (3.23) of all. In the meanwhile RMSE (Root Mean Squared Error) of gas-phase Gibb's free energy changes calculated with DFT (MPW1PW91/6-311+G(2d,2p), G1, G2, G2MP2 and CBS-QB3 methods are 3.49, 4.26, 3.17, 3.12 and 4.61 respectively. In these calculations, G2MP2 method gives a comparatively lower RMSE hence it shows a significant congruence with the experimental values.

Figure 1 shows the correlation of gas phase Gibb's free energy changes calculated by three QSPR methods with experimental values. Among the employed QSPR methods, HYBRID shows a better linearity than others. Therefore HYBRID based QSPR model is more reliable and pertinent methods than other QSPR models for calculation of gas phase Gibb's free energy changes for small molecules. Fig. 2 shows the correlation of Gibb's free energy changes calculated by a DFT (MPW1PW91/6-311+G(2d,2p)) method, Gaussian

n-methods (G1, G2), modified Gaussian method (G2MP2) and a composite method (CBS-QB3) with respect to the experimental values. Upon the analysis of calculations the CBS-QB3 and G1 methods give comparatively poor results, so some calculated values are little apart from the trendline. In the meanwhile, MPW1PW91/6-311+G(2d,2p), G2 and G2MP2 give better results therefore they show a significant improvement of linearity other than two methods (CBS-QB3 and G1).

**Table 2: Experimental and calculated gas-phase Gibb's free energies (kcal/mol) of 74 small molecules in Coral and Gaussian software**

ID	Exp	Graph	SMILES	Hybrid	MPW1PW91 /6-311+G(2d,2p)	G1	G2	G2MP2	CBS-QB3
1	343.20	345.21	322.43	335.86	343.46	342.71	342.73	343.51	341.66
2	347.00	339.86	341.12	341.94	349.81	351.38	351.96	352.21	347.12
3	344.40	344.39	339.69	344.38	346.50	346.29	345.63	346.00	344.78
4	349.00	360.42	349.05	349.07	350.71	345.79	349.09	349.86	349.18
5	345.00	333.63	356.27	339.35	338.43	336.66	336.81	337.30	335.56
6	348.90	345.21	357.45	348.87	350.13	349.04	349.02	349.54	347.71
7	350.60	357.72	356.27	355.64	351.84	351.77	351.63	352.13	349.84
8	327.50	327.51	323.89	327.48	320.61	318.48	320.18	321.09	319.26
9	389.10	378.02	390.07	389.10	391.15	388.86	389.04	388.96	389.81
10	341.00	345.21	333.78	333.95	339.74	340.53	340.72	341.07	339.28
11	347.10	345.57	350.32	347.13	348.69	346.82	347.07	347.60	345.84
12	320.90	324.62	320.73	320.79	320.15	315.92	316.58	317.50	315.31
13	335.50	345.12	338.35	345.39	336.98	334.76	335.86	336.40	309.14
14	344.10	351.68	346.65	344.83	341.82	340.49	341.47	341.51	341.60
15	352.80	344.67	353.71	352.71	358.87	355.12	356.57	356.71	356.59
16	387.90	386.29	391.25	388.03	389.00	387.71	388.56	388.62	387.20
17	395.90	395.91	396.19	395.88	400.62	397.02	396.94	396.67	397.63
18	334.70	344.67	335.62	339.02	332.48	327.15	328.94	329.26	329.08
19	391.70	372.27	387.84	391.30	393.48	390.58	391.17	391.21	390.91
20	355.70	355.75	354.70	355.69	358.24	354.05	355.31	355.46	355.42
21	383.70	383.72	381.31	383.69	387.90	382.59	383.18	383.65	385.80
22	368.50	362.14	369.95	372.79	370.82	367.76	368.96	369.32	369.10
23	369.20	364.83	375.43	372.59	371.84	369.74	370.86	371.12	370.65
24	349.10	360.16	350.33	349.13	350.58	348.61	349.49	349.81	349.79
25	376.20	378.54	373.09	381.24	375.82	374.23	375.45	375.59	371.60
26	369.50	360.68	369.49	369.51	370.45	368.42	368.62	369.59	368.39
27	372.60	365.25	374.26	370.32	372.82	370.56	371.67	371.95	371.65
28	366.60	364.83	362.01	360.36	366.34	365.13	366.20	366.43	366.42
29	367.60	362.40	361.27	374.07	366.75	364.09	364.63	365.42	364.85
30	370.00	396.40	369.64	372.76	371.13	368.52	369.50	369.43	370.35
31	342.50	342.52	356.17	342.61	343.99	341.50	342.88	343.18	343.11
32	360.00	381.17	359.92	357.10	367.56	379.18	364.67	361.84	365.62
33	372.60	392.86	370.81	369.02	374.96	375.99	376.25	375.87	377.08
34	403.00	403.04	402.96	402.97	399.64	417.85	398.10	398.18	398.84
35	352.00	353.67	357.70	357.07	351.44	351.28	351.95	352.45	351.44
36	377.80	389.35	378.02	377.70	380.53	376.91	377.95	377.82	378.36
37	401.00	396.40	394.71	396.25	402.62	400.33	400.96	400.83	401.19
38	364.10	365.71	363.86	364.21	368.04	365.38	365.67	365.49	366.43
39	372.50	371.84	372.62	372.70	374.70	376.04	376.44	376.54	376.68
40	372.80	392.86	371.00	377.81	376.87	375.11	375.38	375.16	376.19
41	357.80	365.25	358.05	354.75	360.83	361.18	362.17	362.19	359.06
42	391.30	397.77	393.79	395.25	393.47	391.62	392.27	392.24	392.63
43	404.00	397.77	397.06	399.81	403.01	398.74	399.16	399.02	398.43
44	391.30	391.33	386.17	391.35	388.06	387.14	388.08	388.07	387.74
45	376.20	397.77	372.17	375.11	379.29	377.71	378.08	377.93	378.23
46	383.80	392.86	395.89	392.51	382.41	383.22	383.31	383.01	383.11
47	379.00	381.17	385.00	380.59	385.30	382.07	383.02	383.37	383.74
48	358.70	364.83	362.50	361.58	358.92	359.39	360.58	360.70	359.04

49	361.90	362.14	357.02	361.78	364.27	362.91	363.60	363.66	363.21
50	398.20	390.01	415.20	398.89	407.03	403.81	404.35	404.49	405.62
51	367.40	367.43	358.52	367.50	367.80	369.03	369.75	369.79	369.36
52	407.20	392.86	409.31	404.73	410.11	409.07	408.92	408.79	408.58
53	411.70	396.40	408.14	408.48	414.25	413.18	413.02	412.92	412.98
54	401.00	389.35	400.83	400.95	407.92	405.33	405.65	405.43	405.62
55	366.40	366.42	366.62	366.38	371.04	368.66	368.47	368.62	371.37
56	344.00	345.57	343.80	343.92	342.95	342.80	343.01	343.39	341.50
57	350.70	350.70	350.57	350.60	351.26	349.61	350.26	350.63	350.73
58	383.00	381.28	372.86	377.95	388.91	387.45	386.95	386.73	386.87
59	383.00	365.25	361.33	359.31	360.36	359.52	360.46	360.47	360.24
60	365.20	372.27	357.34	364.52	365.90	366.88	367.54	367.53	367.51
61	396.50	414.62	379.20	396.74	393.62	391.84	392.20	391.81	393.28
62	394.00	394.04	394.07	394.05	393.50	389.41	390.58	390.37	389.75
63	408.70	408.71	406.96	408.69	412.06	411.15	410.94	410.82	411.67
64	400.60	400.62	397.25	400.64	404.30	402.36	402.70	403.22	403.69
65	381.20	369.48	382.75	381.83	392.44	390.06	390.98	391.73	391.93
66	370.30	370.59	374.94	370.29	372.97	370.79	371.96	372.71	372.65
67	368.00	367.98	363.60	368.02	370.25	369.34	369.85	369.61	369.03
68	349.90	349.92	353.89	349.89	351.73	351.01	352.17	351.74	350.73
69	389.70	378.28	389.88	389.77	386.36	384.61	385.37	385.75	383.77
70	361.30	365.10	348.14	354.14	362.64	361.25	362.42	362.94	362.02
71	342.00	342.01	341.78	342.04	343.08	342.34	344.05	344.62	341.85
72	381.30	381.26	381.23	381.28	386.23	384.85	385.18	384.90	386.50
73	356.00	355.98	356.05	356.01	351.46	349.83	350.07	350.53	351.18
74	356.40	356.40	356.28	356.41	361.28	361.79	361.07	361.72	360.11
RMSE	8.83	5.98	3.23	3.49	4.26	3.17	3.12	4.61	

Upon the statistical analysis of all models, the HYBRID method of CORAL and G2MP2 method of Gaussian 09 gives a better improvement results for gas phase Gibbs free energy changes shown in Table 2. Inspection of these calculated values reveals that in most cases the values are very close to the experimental values. For an instance, in case of compound 9, 21, 31, 51, 57 and 67 the calculated gas-phase Gibb's free energy changes of 389.10, 383.69, 342.61, 367.50, 350.60 and 368.02 kcal/mol by HYBRID method and 388.96, 383.65, 343.18,

369.79, 350.63 and 369.61 kcal/mol by G2MP2 method are very close to the experimental values of 389.10, 383.70, 342.50, 367.40, 350.70 and 368.90 kcal/mol respectively. Now the unsigned errors (%) of these molecules by HYBRID and G2MP2 methods are 0.0, 0.0026, 0.321, 0.0272, 0.0285, 0.2385 and 0.0360, 0.0130, 0.1985, 0.6505, 0.0200, 0.1925 respectively which show a very small deviation compared to experimental values. In addition the unsigned errors of the remaining molecules are almost same or slightly higher from these unsigned errors.

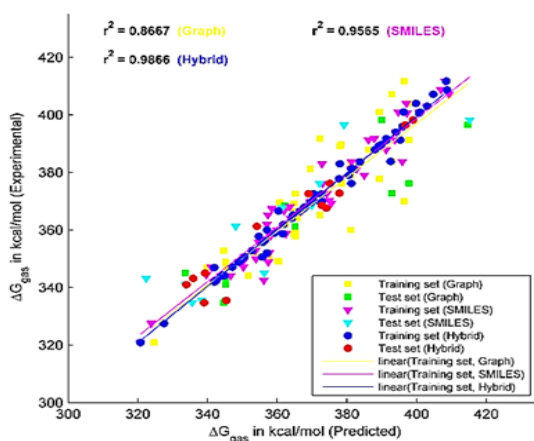


Fig. 1. Comparison between experimental and calculated Gibbs free energies using three QSPR models

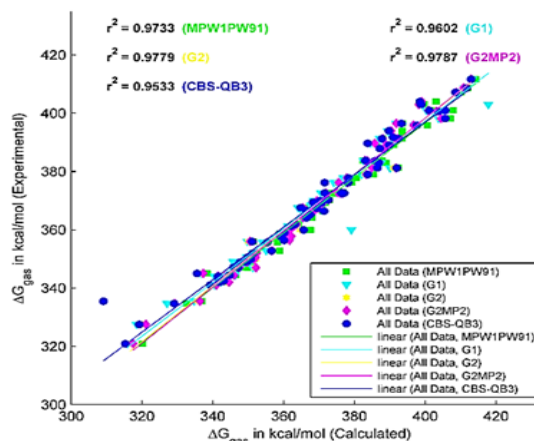


Fig. 2. Comparison between experimental and calculated Gibbs free energies using Gaussian calculations.



**Table 3: Statistical quality of the models for the gas phase Gibb's free energy changes of the training set (n = 61) and the test set (n = 13)**

Statistical Values	Graph		SMILES		Hybrid	
	Training	Test	Training	Test	Training	Test
$r^2$	0.8667	0.8058	0.9565	0.8165	0.9866	0.934
$q^2$	0.8581	0.7381	0.9538	0.7183	0.9857	0.8418
s	7.83	13.2	4.47	11	2.48	5.8
MAE	5.01	10.8	3.09	8.15	1.35	4.83
F	384	46	1296	49	4346	149
$R_m^2$	-	0.6632	-	0.7011	-	0.7382
$\Delta R_m^2$	-	0.1896	-	0.1733	-	0.1081

$r^2$ : squared correlation coefficient  $q^2$ : leave-one-out validation, s: standard error of estimation, MAE: mean -absolute error, F: Fisher ratio

### CONCLUSION

In this paper we have shown gas phase Gibbs free energies of seventy four types of small molecules and these experimentally studied molecules were selected to test new theoretical methodology. There are a variety of methods and tools to calculate the gas phase Gibb's free energies in modern technology. CORAL software is thus one of the modern and an efficient tool to build up a model for Gibb's free energy calculation of a set of diverse substances. The predicted free energy of the applied approach was validated with a test set. The best results were obtained using the HYBRID version of the representation of the molecular structure, i.e., by both molecular graph and SMILES. At the same time Gaussian software program was also employed to calculate the gas phase Gibbs free energies with a variety of methods and basis sets. MPW1PW91/6-311+G(2d,2p) and G2MP2 gave the results with an excellent agreement with experimental values. It is clear that high-levels of theory and cost effective methods are not always feasible for the calculations of gas phase Gibbs free energies of these small

molecules. Sometimes low-levels of theory can reach the point where calculations at the level of chemical accuracy, within 1 kcal/mol, are now possible. The main variation in the accuracy of these molecules stems from the levels of theory used in the optimization. In future, further studies can develop a new methodology to calculate the high accuracy of gas phase Gibbs free energies.

### ACKNOWLEDGEMENT

The authors wish to thank Chairman, Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh to use the departmental computational chemistry laboratory and for his kind support during the work.

### Conflict of Interest

We as authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

### REFERENCES

1. Chipot, C.; Mark, A. E.; Pande, V. S.; Simonson, T. *Springer Ser. Chem. Phys.*, **2007**, *86*, 463–492.
2. Kollman, P. *Chem. Rev.*, **1993**, *93*, 2395–2417.
3. Ateshian, G. A.; Morrison, B.; Hung, C. T.; Hung, C. T. *Ann. Biomed. Eng.*, **2010**, *38*(5), 1801–1814.
4. Scholtz, J. M.; Pace, C. N.; Huyghues-Despointes, B. M. P. *Nat. Struct. Biol.*, **1999**, *6*(10), 910–912.
5. Vaes, W. H. J.; Ramos, E. U.; Verhaar, H. J. M.; Christopher J. Cramer; Hermens, J. L. M. *Chem. Res. Toxicol.*, **1998**, *11*(8), 847–854.
6. Lozano, H. R.; Martínez, F. *Brazilian J. Pharm. Sci.*, **2006**, *42*(4), 601–613.
7. Van Den Bosch, M.; Swart, M.; Snijders, J. G.; Berendsen, H. J. C.; Mark, A. E.; Oostenbrink, C.; Van Gunsteren, W. F.; Canters, G. W. *Chem Bio Chem.*, **2005**, *6*, 738–746.



8. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.*, **1998**, *120*, 9401–9409.
9. Chipot, C. *In New Algorithms for Macromolecular Simulation*; Springer-Verlag: Berlin/Heidelberg., **2006**, 185–211.
10. Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.*, **2009**, *31*(8), NA-NA.
11. Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; *Comput. Phys. Commun.*, **2009**, *180* (10), 1961–1972.
12. Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.*, **2003**, *330* (4), 891–913.
13. Koca, J.; Jirouskova, Z.; Svobodova Varekova, R.; Vanek, J. *J. Comput. Chem.*, **2009**, *30*(7), 1174–1178.
14. Aledo, J. C. *Biochem. Mol. Biol. Educ.*, **2001**, *29*(4), 142–143.
15. Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.*, **1982**, *17*, 55–67.
16. Warshel, A. *J. Phys. Chem.*, **1982**, *86*, 2218–2224.
17. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* (80-.), **1987**, *236*, 564–568.
18. Bash, P. A.; Chandra Singh, U.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* (80-.), **1987**, *235* (4788), 574–576.
19. Westacott, r. E.; Rodger, P. M. *Ann. N. Y. Acad. Sci.*, **1994**, *715* (1), 539–543.
20. CORAL/Free software for QSAR and nanoQSAR, <http://www.insilico.eu/CORAL/> (accessed Apr 14., **2019**).
21. Toropova, A. P.; Toropov, A. A.; Martyanov, S. E.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. *Chemom. Intell. Lab. Syst. Short Commun.*, **2012**, *110*(1), 177–181.
22. Robinson, M. S.; Davico, G. E.; Bierbaum, V. M.; DePuy, C. H. *Int. J. Mass Spectrom. Ion Process.*, **1994**, *137*(C), 107–119.
23. O'Hair, R. A. J.; DePuy, C. H.; Bierbaum, V. M. *J. Phys. Chem.*, **1993**, *97*, 7955–7961.
24. Rempala, K.; Ervin, K. M. *J. Chem. Phys.*, **2000**, *112* (10), 4579–4590.
25. Gilles, M. K.; Polak, M. L.; Lineberger, W. C. *J. Chem. Phys.*, **1992**, *96* (11), 8012–8020.
26. Downard, K. M.; Bowie, J. H.; O'Hair, R. A. J.; Krempp, M.; DePuy, C. H. *Int. J. Mass Spectrom. Ion Process.*, **1992**, *120*(3), 217–229.
27. Bartmess, J. E.; Scott, J. A.; McIver, Robert T., J. *J. Am. Chem. Soc.*, **1979**, *101*(20), 6046–6056.
28. Burk, P.; Koppel, I. A.; Rummel, A.; Trummel, A. *J. Phys. Chem., A* **2000**, *104*(7), 1602–1607.
29. Ingemann, S.; Nibbering, N. M. M. *Compd. J. Chem. Soc. Perkin Trans.*, **2014**, *2*, 837.
30. Guillemin, J. C.; Rague, E. H.; Gal, J. F.; Maria, P. C.; Mó, O.; Yáñez, M. *Chem. Eur. J.*, **2005**, *11*(7), 2145–2153.
31. Bierbaum, V. M.; Grabowski, J. J.; DePuy, C. H. *J. Phys. Chem.*, **1984**, *88* (7), 1389–1393.
32. Yang, X.; Kiran, B.; Wang, X.; Wang, L.; Mucha, M.; Jungwirth, P. *J. Phys. Chem., A* **2004**, *108*, 7820–7826.
33. Koppel, I. A. ; Taft, R. W. ; Anvia, F. ; Zhu, S. Z.; Hu, L. Q. ; Sung, K. S. ; Desmarteau, D. D.; Yagupolskii, L. M.; Yagupolskii, Y. L.; Ignat'ev, N. V.; *J. Am. Chem. Soc.*, **1994**, *116*(7), 3047–3057.
34. Taft, R. W. *Progress in Physical Organic Chemistry*, John Wiley & Sons: New York., **1987**, *16*.
35. Radisic, D.; Xu, S.; Bowen, K. H. *J. Chem. Phys. Lett.*, **2002**, *354*, 9–13.
36. Wickham-Jones, C. T.; Ervin, K. M.; Ellison, G. B.; Lineberger, W. C. *J. Chem. Phys.*, **1989**, *91*(4), 2762–2763.
37. Bradforth, S. E.; Kim, E. H.; Arnold, D. W.; Neumark, D. M. *J. Chem. Phys.*, **1993**, *98*(2), 800–810.
38. Mackay, G. I.; Hems Worth, R. S.; Bohme, D. K. *Can. J. Chem.*, **1976**, *54*, 1624.
39. Ma, S.; Wang, F.; Cooks, R. G. *J. Mass Spectrom.*, **1998**, *33*(10), 943–949.
40. Smith, J. R.; Kim, J. B.; Lineberger, W. C. *Phys. Rev., A* **1997**, *55*(3), 2036–2043.
41. Ramond, T. M.; Davico, G. E.; Schwartz, R. L.; Lineberger, W. C. *J. Chem. Phys.*, **2000**, *112*(3), 1158–1169.
42. Ellison, G. B.; Engelking, P. C.; Lineberger, W. C. *J. Phys. Chem.*, **1982**, *86*(25), 4873–4878.
43. Distelrath, V.; Boesl, U. *Faraday Discuss.*, **2000**, *115*(1), 161–174.
44. Nee, M. J.; Osterwalder, A.; Zhou, J.; Neumark, D. M. *J. Chem. Phys.*, **2006**, *125*, 014306(1-9).

45. Ramond, T. M.; Blanksby, S. J.; Kato, S.; Bierbaum, V. M.; Davico, G. E.; Schwartz, R. L.; Lineberger, W. C.; Ellison, G. B. *J. Phys. Chem.*, A **2002**, *106*(42), 9641–9647.
46. Graul, S. T.; Schnute, M. E.; Squires, R. R. *Int. J. Mass Spectrom. Ion Process.*, **1990**, *96*, 181–198.
47. Blanksby, S. J.; Ramond, T. M.; Davico, G. E.; Nimlos, M. R.; Kato, S.; Bierbaum, V. M.; Lineberger, W. C.; Ellison, G. B.; Okumura, M. *J. Am. Chem. Soc.*, **2001**, *123*(39), 9585–9596.
48. Botter, R.; Dibeler, V. H.; Walker, J. A.; Rosenstock, H. M. *J. Chem. Phys.*, **1966**, *44* (3), 1271–1278.
49. Akin, F. A.; Ervin, K. M. *J. Phys. Chem.*, A **2006**, *110*(4), 1342–1349.
50. Thynne, J. C. ; MacNeil, K. A. G. *J. Phys. Chem.*, **1971**, *75* (17), 2584–2591.
51. Gal, J. F.; Decouzon, M.; Maria, P. C.; González, A. I.; Mó, O.; Yáñez, M.; El Chaouch, S.; Guillemin, J. C. *J. Am. Chem. Soc.*, **2001**, *123* (26), 6353–6359.
52. Graul, S. T.; Squires, R. R. *J. Am. Chem. Soc.*, **1990**, *112*(7), 2517–2529.
53. Dorofeeva, O. V. *J. Phys. Chem. Ref. Data.*, **1986**, *15*, 437–464.
54. Ervin, K.; Gronert, S.; Barlow, S.; Gilles, M.; Harrison, A.; Bierbaum, V. M.; Depuy, C. H.; Lineberger, W. C. *J. Am. Chem. Soc.*, **1990**, *112* (1), 5750–5759.
55. Clifford, E. P.; Wenthold, P. G.; Lineberger, W. C.; Petersson, G. A.; Broadus, K. M.; Kass, S. R.; Kato, S.; DePuy, C. H.; Bierbaum, V. M.; Ellison, G. B. *J. Phys. Chem.*, A **1998**, *102* (36), 7100–7112.
56. Bartmess, J. E.; Burnham, R. D. *J. Org. Chem.*, **1984**, *49*(8), 1382–1387.
57. Robinson, M. S.; Bierbaum, V. M.; DePuy, C. H.; Polak, M. L.; Lineberger, W. C. *J. Am. Chem. Soc.*, **1995**, *117*(25), 6766–6778.
58. Oakes, J. M.; Jones, M. E.; Bierbaum, V. M.; Ellison, G. B. *J. Phys. Chem.*, **1983**, *87*, 4810–4815.
59. de Visser, S. P.; de Koning, L. J.; van der Hart, W. J.; Nibbering, N. M. M. *Recl. des Trav. Chim. des Pays Bas.*, **1995**, *114* (6), 267–272.
60. DePuy, C. H.; Gronert, S.; Barlow, S. E.; Bierbaum, V. M.; Damrauer, R. *J. Am. Chem. Soc.*, **1989**, *111*, 1968–1973.
61. Dahlke, G. D.; Kass, S. R. *J. Am. Chem. Soc.*, **1991**, *113* (15), 5566–5573.
62. Parr, A. C. ; Elder, F. A. *J. Chem. Phys.*, **1968**, *49*, 2659.
63. Ellison, G. B.; Davico, G. E.; Bierbaum, V. M.; DePuy, C. H. *Int. J. Mass Spectrom. Ion Process.*, **1996**, *156* (1–2), 109–131.
64. Rabasco, J. J.; Kass, S. R. *J. Am. Soc. Mass Spectrom. Soc. Mass Spectrom.*, **1992**, *3*, 91–98.
65. DePuy, C. H.; Bierbaum, V. M.; Damrauer, R. *J. Am. Chem. Soc.*, **1984**, *106* (8), 4051–4053.
66. Plemenkov, V. V. ; Villem, Y. Y. ; Villem, N. V. ; Bolesov, I. G. ; Surmina, L. S.; Yakushkina, N. I. ; Formanovskii, A. A. *Zh. Obs. Khim.*, **1981**, *51*, 2076.
67. Zhang, L.; Grabowski, J. J. *J. Chem. Soc., Chem. Commun.*, **1989**, *1*, 1819–1821.
68. Moran, S.; Barney Ellison, G. *J. Phys. Chem.*, **1988**, *92* (7), 1794–1803.
69. Tian, Z.; Fattahi, A.; Lis, L.; Kass, S. R. *J. Am. Chem. Soc.*, **2006**, *128*(51), 17087–17092.
70. Kroeker, R. L.; Kass, S. R. *J. Am. Chem. Soc.*, **1990**, *112*, 9024–9025.
71. Ellison, G. B.; Engelking, P. C.; Lineberger, W. C. *J. Am. Chem. Soc.*, **1978**, *100*(8), 2556–2558.
72. Deyerl, H.-J.; Alconcel, L. S.; Continetti, R. E. *J. Phys. Chem.*, A **2001**, *105*, 552–557.
73. Born, M.; Ingemann, S.; Nibbering, N. M. M. *Int. J. Mass Spectrom.*, **2000**, *194* (2–3), 103–113.
74. Paulino, J. A.; Squires, R. R. *J. Am. Chem. Soc.*, **1991**, *113* (15), 5573–5580.
75. Hierl, P. M.; Henschman, M. J.; Paulson, J. F. *Int. J. Mass Spectrom. Ion Process.*, **1992**, *117* (C), 475–485.
76. Rackwitz, R.; Feldmann, D.; Kaiser, H. ; Heinicke, E. Z. *Naturforsch.*, A **1977**, *32*, 594–599.
77. Bengali, A. A.; Leopold, D. G. *J. Am. Chem. Soc.*, **1992**, *114*, 9192–9194.
78. Ervin, K. M.; Lineberger, W. C. *J. Chem. Phys.*, **2005**, *122* (19), 1–11.
79. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Gaussian Inc., Wallingford CT., **2009**.

80. Roothaan, C. C. *J. Rev. Mod. Phys.*, **1951**, *23* (2), 69–89.
81. Pople, J. A.; Nesbet, R. K. *J. Chem. Phys.*, **1954**, *22*, 571–572.
82. McWeeny, R.; Diercksen, G. *J. Chem. Phys.*, **1968**, *49*, 4852.
83. Pople, J. A.; Head Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.*, **1989**, *90* (10), 5622–5629.
84. Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.*, **1990**, *93* (4), 2537–2545.
85. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.*, **1991**, *94*(11), 7221–7230.
86. Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.*, **1993**, *98*(2), 1293–1298.
87. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.*, **1999**, *110* (6), 2822–2827.
88. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.*, **2000**, *112* (15), 6532–6542.
89. Adamo, C.; Barone, V. *J. Chem. Phys.*, **1998**, *108* (2), 664–675.
90. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.*, **1999**, *110* (6), 2822–2827.
91. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.*, **1998**, *109* (18), 7764–7776.